

**Woran bemisst sich eine "gute" allgemeine
Bevölkerungsumfrage? Analysen zu Ausmaß,
Bedeutung und zu den Hintergründen
von Nonresponse in zufallsbasierten
Stichprobenerhebungen am Beispiel des ALLBUS**

Schneekloth, Ulrich; Leven, Ingo

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schneekloth, U., & Leven, I. (2003). Woran bemisst sich eine "gute" allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. *ZUMA Nachrichten*, 27(53), 16-57. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-207728>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

WORAN BEMISST SICH EINE „GUTE“ ALLGEMEINE BEVÖLKERUNGSUMFRAGE?

Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS

ULRICH SCHNEEKLOTH & INGO LEVEN

Gemeinhin gilt die Ausschöpfung als eines der wichtigsten Merkmale zur Beurteilung der Qualität von repräsentativen allgemeinen Bevölkerungsumfragen. Anhand der Ergebnisse der „best practice“ ALLBUS-Erhebungen der Jahre 1994, 1996 und 2000 lässt sich zeigen, dass die üblicherweise geforderten hohen Ausschöpfungsraten bei harter Dokumentation der Bearbeitung faktisch nicht realisierbar sind. In der Arbeit wird der Frage nachgegangen, inwieweit eine gemeinhin als „schlecht“ angesehene Ausschöpfung auch tatsächlich zu einer mangelhaften Datenqualität führt. Dabei werden drei unterschiedliche Ansätze zur Beurteilung der Ergebnisqualität exemplarisch vorgeführt. In einem ersten Schritt wird ein Abgleich einiger soziodemografischer Randverteilungen der Stichprobe mit dem deutschen Mikrozensus als Referenzstatistik (Nettovalidierung) vorgenommen. In einem zweiten Schritt wird eine (multivariate) Nonresponse-Analyse durchgeführt. In einem dritten Schritt werden schließlich die Effekte der beim ALLBUS üblichen umfangreichen Nachbearbeitung von Adressen, für die bis dahin noch kein Interview realisiert werden konnte, dargestellt. Die Analysen zeigen, dass zufallsbasierte Stichproben auch bei einer Ausschöpfung von weniger als 50 Prozent keine nennenswerten systematischen Verzerrungen aufweisen müssen. Die isolierte Betrachtung der Ausschöpfung erweist sich von daher als ungeeigneter Prädiktor für Ergebnisqualität.

Sample response rates are usually considered to be one of the most important indicators of quality for representative general population surveys. The analysis of the "best practice" ALLBUS surveys of 1994, 1996 and 2000 shows that the high response rates typically called for in this kind of survey stand in conflict with strict documentation requirements for the fielding process. The article discusses whether so-called "poor" response

rates really impair data quality. Three different examples are presented for approaching the problem of judging the quality of random samples. First, we adjust some socio-demographic marginal totals of the random probability sample to those of the German micro-census as a reference statistic (Nettovalidation). Second, we carry out a (multivariate) non-response-analysis. Finally, we present the effects of the post-processing (re-working) of addresses usually employed in ALLBUS surveys for addresses for which interviews have not yet been realized. Our analysis shows that random samples are not necessarily systematically biased, even if the sample response rate is less than 50%. Focusing solely on the sample response rate thus proves to be an inappropriate indicator for data quality.

1. Problemstellung

Repräsentativ angelegte Erhebungen stellen überall dort, wo zum Zwecke der Analyse auf ansonsten nicht verfügbare Daten zurückgegriffen werden muss, den „Königsweg“ der Forschung dar.¹ Im Unterschied zu sonstigen Stichprobenverfahren, die entweder ungeregelt, nach dem Schneeballverfahren oder zum Beispiel auf Basis von vorab definierten Quotenvorgaben organisiert werden, basieren diese Verfahrensmodelle darauf, dass die Auswahl der Zielpersonen streng nach Zufallskriterien vorgenommen wird.² Nur dann, wenn die Auswahlchancen für alle Zielpersonen gleich oder zumindest mathematisch berechenbar sind, ist es zulässig, die Ergebnisse einer Stichprobe anhand von definierten statistischen Regeln als repräsentativ für die Grundgesamtheit zu betrachten.

Repräsentativerhebungen kommen heute in der ganzen Breite der Sozial-, Markt- und Meinungsforschung zum Einsatz. Neben den einschlägigen wissenschaftlichen Erhebungen mit Modellcharakter, wie zum Beispiel dem Sozio-ökonomischen Panel (SOEP) oder der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS), gilt dies auch für einen nicht unerheblichen Teil von Datenbeständen, die offiziell unter dem Titel „Amtliche Statistik“ geführt werden. Verwiesen werden kann in diesem Zusammenhang auf die im jährlichen Turnus erhobene Mikrozensus-Stichprobe des Statistischen Bundes-

1 Nicht umsonst stellt auch die 1999 von Max Kaase herausgegebenen Denkschrift der Deutschen Forschungsgemeinschaft „Qualitätskriterien der Umfrageforschung“ fest dass „Repräsentative Umfragen (...) sich seit dem Zweiten Weltkrieg zum unentbehrlichen Mittel der Datenerhebung über größere Bevölkerungen entwickelt“ haben. (DFG 1999: 12).

2 Ein Überblick zu den gängigen Auswahlverfahren findet sich zum Beispiel bei Schnell/Hill/ Esser (1999: 247ff.)

amtes oder auch seit neuestem auf die von Infratest Sozialforschung für das Statistische Bundesamt durchgeführte Erhebung „Der Arbeitsmarkt in Deutschland“.³

Umso bemerkenswerter ist es, dass Fragestellungen, die sich auf die Qualität von Umfragen beziehen, im wissenschaftlichen Diskurs nach wie vor eher „stiefmütterlich“ behandelt werden und etwas überspritzt formuliert als „Steckenpferd“ eines eher kleineren Zirkels von Methodenforscherinnen und -forschern erscheinen. Aus unserer Sicht gibt es in Deutschland heutzutage nach wie vor keinen wirklich praxistauglichen Konsens darüber, welche Anforderungen an eine definierte Qualität von Umfragen zu stellen sind (Kriterien) bzw. mit welchen Messkonzepten Qualität möglichst kontinuierlich auch empirisch überprüft werden kann (Monitoring). Unser Eindruck ist der, dass nach wie vor zwei Extremstrategien im Rahmen einer „Qualitätssicherung“ in der empirischen Sozialforschung vorherrschen. Entweder man lässt die Dinge laufen und verlässt sich darauf, dass die „durchführenden Instanzen“ (Umfrageinstitute, sonstige private Anbieter, Lehrstühle, Institute etc.) die Dinge schon richtig machen werden, oder aber man formuliert zum Beispiel im Rahmen der Ausschreibung – unabhängig vom Studententyp oder dem inhaltlichen Erkenntnisinteresse – in der Regel „Lehrbuch gestützt“ alle möglichen Maximalanforderungen, um sich dann hinterher doch für das Angebot zu entscheiden, das in einem als finanziell noch verkraftbar betrachteten Rahmen kalkuliert ist. Beide Strategien sind aus unserer Sicht nicht per se zielführend oder qualitätsfördernd. Das eigentliche Dilemma besteht darin, dass es jenseits möglicher Ankündigungen, welche methodischen Standards im Einzelnen sinnvoll sein könnten bzw. welche auch tatsächlich praxistauglich und/oder finanzierbar sind, für den Endnutzer nach wie vor viel zu wenig Kriterien für Ergebnisqualität gibt, mit denen sich die Belastbarkeit der erhobenen Daten einigermaßen verlässlich abschätzen lässt.

1.1 Zur Qualität von Umfragen

Die Denkschrift der Deutschen Forschungsgemeinschaft zu den „Qualitätskriterien der Umfrageforschung“ führt hierzu aus, dass es jenseits der theoretischen Kriterien einer verzerrungsfreien Abbildung der Grundgesamtheit durch die Stichprobe sowie der zuverlässigen Messung der Sachverhalte anhand valider Erhebungsinstrumente praktisch zwei Arten von Indizien gibt, um die Qualität von Umfragen zu bewerten:

- Richtige Ergebnisse sowie
- korrekte Verfahren (DFG 1999: 94).

3 Vergleiche dazu die entsprechenden Darstellungen auf der Web-Site des Statistischen Bundesamtes unter http://www.destatis.de/dt_erheb/arbeitsmarkt/am_einf.htm bzw. zum Mikrozensus unter http://www.destatis.de/micro/d/micro_c1.htm

Die Kriterien mögen auf den ersten Blick trivial erscheinen. Tatsächlich markieren sie aber die beiden zentralen Eckpunkte, an denen sich Umfragen bemessen lassen müssen.

Richtige Ergebnisse sind selbstredend das zentrale Ziel von Forschung. Überall dort, wo keine einfachen Tatbestände, die in der Wirklichkeit mehr oder weniger leicht quantifizierbar sind, sondern komplexere Strukturen gemessen werden, kann man dem erhobenen Ergebnis allerdings nicht immer unmittelbar „ansehen, ob und bis zu welchem Grade es denn „richtig“ oder „falsch“ ist. Häufig hilft hier nur das Vertrauen in die „Regeln der Kunst“, wie es weiter in der zitierten Denkschrift heißt, sprich in die korrekte Anlage und Durchführung der Erhebung. Wir stimmen dieser Einschätzung zu und möchten hierzu ergänzend anmerken, dass es zusätzlich darauf ankommt, für beide vorhin benannten Kriterien zu praktikablen Operationalisierungen zu kommen, die helfen, sowohl ex ante als auch ex post die Qualität einer Umfrage zu beurteilen.

Folgen möchten wir auch der weiteren Argumentation der Denkschrift:

„Diese Vertrauensbildung (in die Belastbarkeit der Ergebnisse – US/IL) über Verfahren ist wichtig und unverzichtbar, aber sie hat ihre Risiken. Bedenklich ist etwa, wenn einzelne Kriterien als Qualitätsmaßstab verabsolutiert werden (beispielsweise eine Stichprobenausschöpfung in einer bestimmten Höhe. (...) Problematisch wird es auch dann, wenn die Orientierung an etablierten Verfahren zu einem methodischen Konservativismus führt.“ (DFG 1999: 94f)

In der Praxis wird die Güte einer Erhebung häufig isoliert anhand der realisierten **Ausschöpfung** beurteilt. Unter Ausschöpfung wird dabei in der Regel das Verhältnis der realisierten Netto-Interviews zu den ursprünglich im Rahmen der Stichprobe vorgegebenen gültigen Bruttofällen verstanden. Der daraus resultierende Quotient hat ohne Zweifel seinen Charme. Drückt man das Ergebnis in Prozent, also im Verhältnis zum Idealwert 100 aus, so ergibt sich eine lineare Größe, der man scheinbar auf den ersten Blick ansieht, ob sie hoch oder niedrig, akzeptabel oder schlecht ist. Was liegt näher, als die Qualität einer Umfrage anhand dieses linearen Kriteriums im Sinne eines Benchmarkings zu messen.

Tatsächlich misst die Ausschöpfung jedoch nur – worauf ebenfalls völlig zu Recht in der Denkschrift der DFG hingewiesen wird – „wie groß der Spielraum für Selektivität durch Nonresponse ist. Sie besagt nichts über die tatsächliche Selektivität.“ (DFG 1999: 104) Nonresponse führt nur dann zu einer Verschlechterung der Erwartungstreue, wenn damit systematisch eine Unterausschöpfung von bestimmten Bevölkerungsgruppen verbunden ist. Ein Nonresponse, der mehr oder weniger zufällig verteilt ist, ist hingegen unter Qualitäts Gesichtspunkten betrachtet unerheblich. Die Effekte von Nonresponse müssen dem-

nach erst im Einzelnen überprüft werden, bevor man eine mindere Qualität postuliert. Unter anderem aus diesem Grunde gibt es in der Literatur bisher auch kein allgemeingültiges Kriterium, wie hoch eine Ausschöpfung ausfallen muss, um eine zufallsbasierte Stichprobe als adäquate Abbildung der Grundgesamtheit zu charakterisieren. Eine formale Regel, die besagt, dass eine Stichprobe zum Beispiel zu mindestens 70 Prozent ausgeschöpft sein soll, um Anerkennung zu finden, lässt sich letztendlich weder theoretisch noch empirisch herleiten.⁴

1.2 Empirischer Zugang

Wir möchten mit der vorliegenden Arbeit einen Beitrag dazu leisten, die tatsächlich erreichbaren Anforderungen an die Qualität von Umfragen empirisch zu beleuchten. Einen wichtigen Stellenwert nimmt dabei die Frage ein, inwieweit die Höhe der Ausschöpfung für sich genommen bereits einen hinreichend geeigneten Indikator zur Beurteilung der Qualität einer repräsentativ angelegten Erhebung darstellt. Oder anders formuliert: Kann es gute Qualität auch bei „schlecht“ ausgeschöpften Stichproben geben?

Als empirische Basis haben wir hierfür die von Infratest Sozialforschung zuletzt im Jahr 2000 im Auftrag von ZUMA, Mannheim durchgeführten ALLBUS-Erhebungen herangezogen.⁵ Der ALLBUS eignet sich aus den folgenden Gründen gut für eine methodische Reanalyse.

Zum Ersten handelt es sich beim ALLBUS dem eigenen Anspruch nach um ein „best practice Verfahrensmodell“, bei dem im zweijährigen Turnus persönlich-mündliche Befragungen auf der Basis von Zufallsstichproben durchgeführt werden. Die Stichprobenziehung erfolgte in den Jahren 1994, 1996 und 2000 anhand einer Einwohnermeldeamtstichprobe (Registerstichprobe). Im Unterschied zum weit verbreiteten und alternativen Random-Route Modell, bei dem der Interviewer ausgehend von einer vorgegebenen und

4 Bei den im Rahmen der scientific community geforderten Schwellenwerten handelt es sich aus unserer Sicht um Faustformeln, die aus den vorhandenen Informationen über früher üblicherweise dokumentierte Ausschöpfungen abgeleitet werden. Schnell, Hill, Esser weisen zum Beispiel aus, dass die „typischen“ Ausschöpfungen in den 70iger Jahren bei 80 Prozent lagen, während seit Beginn der 80iger Jahre eher Werte um die 70 Prozent üblich seien (Schnell/Hill/Esser 1999: 288).

5 Die ALLBUS-Stichproben der Jahre 1994, 1996 und 2000 wurden von Infratest Sozialforschung als Registerstichprobe nach dem jeweils gleichen Verfahrensmodell durchgeführt. Die Daten liegen vor und können über das Zentralarchiv für empirische Sozialforschung (ZA) zum Zwecke wissenschaftlicher Analysen angefordert werden. Die Erhebung von 1998 unterscheidet sich methodisch und wurde von GFM-Getas als Random-Route Befragung durchgeführt. Die Ergebnisse der neuesten Erhebung von 2002, die von Infas ebenfalls auf Basis einer Einwohnermeldeamtstichprobe erstellt wurde, lagen uns noch nicht vor.

zufällig gezogenen Startadresse anhand einer festen Begehungsregel zum Beispiel jeden dritten Haushalt kontaktieren soll (Random-Walk), werden die Adressen in diesem Fall anhand einer ebenfalls zufällig gezogenen Auswahl von Einwohnermeldeämtern bestimmt und danach dem Interviewer vorab fest vorgegeben.⁶ Stichprobenziehung und Ergebnisse der Feldbearbeitung sind in diesem Fall eindeutig getrennt und als mögliche Nettoeffekte im Vergleich zur Bruttostichprobe klar abgrenzbar. Die Feldarbeit erfolgt mit großer Intensität und ist in ihren zentralen Bestandteilen gegenüber dem Auftraggeber transparent und nachvollziehbar angelegt. Die „Regelgebundenheit“ in der Bearbeitung ist damit genuiner Bestandteil des Verfahrensmodells.⁷

Anlage und Prozess der Feldarbeit sind **zum Zweiten** bei den ALLBUS-Untersuchungen umfangreich dokumentiert.⁸ Ein methodisches Kernstück des Verfahrensmodells der Jahre 1994, 1996 und 2000 (Adressen-Random bzw. AR-Modell) bildet die Bereitstellung eines sog. Bruttobandes, in dem alle im Rahmen der Erhebung brutto bei den Einwohnermeldeämtern gezogenen Adresseneinheiten (anonymisiert) aufgeführt und hinsichtlich ihres jeweiligen Bearbeitungsergebnisses nachgewiesen sind. Die vorliegende Messreihe der Jahre 1994, 1996 und 2000 erlaubt von daher einen tatsächlichen Vergleich der Entwicklung der Ausschöpfung, da in diesem Fall die Berechnung der brutto eingesetzten und netto auch tatsächlich realisierten Adressen nach dem gleichen Prinzip erfolgt.

Gerade dieser letzte Aspekt ist unter methodischen Gesichtspunkten zentral. In der Praxis können Ausschöpfungen bei unterschiedlichen Verfahrensmodellen nicht völlig vergleichbar berechnet werden. Tatsächlich macht es einen bedeutenden Unterschied aus, ob die Bruttostichprobe vorab definiert und vorgegeben ist, oder ob der Interviewer zum Beispiel im Rahmen eines Random-Walks im gleichen Schritt aufschreibt, wie viele Haushalte er brutto kontaktiert und wie viele er davon tatsächlich netto befragen konnte (Random-Route Verfahren). Abweichungen bei der Dokumentation von Fehlkontakten können zu deutlichen Veränderungen bezüglich der errechneten Ausschöpfung führen.⁹

6 Hierzu wird im ersten Schritt eine nach regionalen und siedlungsstrukturellen Kriterien geschichtete Zufallsstichprobe von Gemeinden erstellt. Im Anschluss daran werden die so genannten Sample-Points – eine größenproportional zur ausgewählten Gemeinde vorgegebene Menge von Zielpersonenadressen – anhand einer Allokationsrechnung ermittelt, die dann in einem zweiten Schritt von den jeweiligen Einwohnermeldeämtern der Gemeinden ebenfalls zufällig aus dem Melderegister gezogen werden. Vergleiche dazu zum Beispiel Koch/Wasmer/Harkness/Scholz (2001).

7 Vergleiche dazu im Einzelnen die Kriterien für Stichprobenqualität in DFG (1999: 100f).

8 Die von ZUMA erstellten Methodenberichte zu den ALLBUS-Untersuchungen stehen im Internet unter <http://www.gesis.org/Dauerbeobachtung/ALLBUS/Bestellen/mberichte.htm> zum Download bereit.

9 Hierbei muss es sich übrigens nicht immer um bewusste Fälschungen handeln. Gerade beim Random-Walk ist es für den Interviewer häufig schwierig zu entscheiden, ob sich hinter einer

Verweisen möchten wir an dieser Stelle auf eine Arbeit von Koch aus dem Jahre 1998, der anhand eines Vergleiches verschiedener allgemeiner Bevölkerungsstichproben zu dem Ergebnis kommt, das ausgewiesenes „mehr“ nicht gleichbedeutend mit „besser“ sein muss (Koch 1998). Hierbei zeigte sich, dass für „hart“ dokumentierte Registerstichproben, wie etwa dem ALLBUS 1994, deutlich niedrigere Ergebnisse im Bereich um die 55 Prozent als für vergleichbare Random-Route Erhebungen mit Ausschöpfungen um die 70 Prozent (Sozialwissenschaften-Bus) bzw. sogar bis zu 80 Prozent (Media-Analyse) angegeben wurden. Ein Vergleich der Randverteilungen der Stichproben mit den Daten des Mikrozensus ergab jedoch, dass die scheinbar „schlechtere“ Registerstichprobe alles in allem betrachtet die beste Übereinstimmung mit der externen Referenzstatistik aufwies.

Koch merkt in diesem Zusammenhang an, dass

„zu fragen ist ..., inwieweit die berichteten Ausschöpfungsergebnisse – insbesondere die hohen Werte von 70 Prozent und mehr – die Realität korrekt wiedergeben. (...) Bei den analysierten Umfragen kann lediglich im ALLBUS 1994 die Ausschöpfung (weitgehend) zuverlässig berechnet werden, weil aufgrund der Ziehung einer Stichprobe von Personenadressen aus Einwohnermelderegistern vor Feldbeginn ein genau definiertes Stichprobenbrutto vorlag (...). Sämtliche andere Umfragen dagegen wurden nach dem ADM-Design durchgeführt. Die faktische Bruttostichprobe wird hier durch die Interviewer in der zweiten und dritten Auswahlstufe gebildet. Die Zuverlässigkeit der Ausschöpfungsbeurteilung hängt davon ab, wie exakt die Interviewer die Vorgaben betrachten.“ (Koch 1998: 81).

Diese Hypothese wird auch durch neuere Befunde zur tatsächlich erreichbaren Ausschöpfung bei allgemeinen Bevölkerungsstichproben erhärtet. So wurden etwa im Rahmen des DEFECT-Projekts unter methodisch kontrollierten Bedingungen bei face-to-face Befragungen mit fest vorgegebenen Haushaltsadressen Ausschöpfungen von 39,3 bzw. 41,5 Prozent berichtet. Hierbei wurden die Adressen von Projektmitarbeiterinnen und -mitarbeitern auf Basis eines Random-Walks aufgeschrieben und den einbezogenen kommerziellen Umfrageinstituten extern vorgegeben. Für eine vergleichbar angelegte Telefon-

Adresse ein Privathaushalt oder zum Beispiel eine Geschäftsadresse verbirgt, ob der Haushalt, bei dem auch bei mehrmaligem Anlauf kein Kontakt herzustellen ist, grundsätzlich unbewohnt ist oder ob dort eine zur Grundgesamtheit gehörige Person lebt, die schwer zu erreichen ist. Dokumentiert der Interviewer die Adresse als nicht zur Grundgesamtheit zugehörig, so gilt der Ausfall als qualitätsneutral und wird bei der Berechnung der Ausschöpfung nicht mit berücksichtigt. Dokumentiert er hingegen einen Ausfallcode wie zum Beispiel „im Haushalt niemanden angetroffen“, so sinkt entsprechend die Ausschöpfung.

erhebung wurde eine Ausschöpfung von 29,4 Prozent und für eine schriftliche Befragung eine Rücklaufquote von 48,7 Prozent ermittelt (Schnell/Kreuter 2000: 96).

Auch die Ausschöpfungen der ALLBUS-Registerstichproben der Jahre 1994 und 1996 haben im Bereich zwischen 53 Prozent und 55 Prozent gelegen – Tendenz fallend. Beim ALLBUS 2000 wurde sogar zum ersten Mal die „50 Prozent-Marke“ unterschritten (vergleiche dazu Übersicht 1), was offenbar von der einschlägigen scientific community als „Alarmsignal“ interpretiert wurde.

Übersicht 1: Anlage und Ergebnisse der register-basierten ALLBUS-Untersuchungen

	ALLBUS 2000		ALLBUS 1996		ALLBUS 1994	
	West	Ost	West	Ost	West	Ost
Grundgesamtheit	Bevölkerung ab 18 Jahren in Privathaushalten in Deutschland		Bevölkerung ab 18 Jahren in Privathaushalten in Deutschland		Bevölkerung ab 18 Jahren in Privathaushalten in Deutschland	
Stichprobe	Einwohnermeldeamtsstichprobe		Einwohnermeldeamtsstichprobe		Einwohnermeldeamtsstichprobe	
Erhebungsmethode	persönlich/mündlich		persönlich/mündlich		persönlich/mündlich	
Geplante durchschnittliche Interviewdauer	50 Minuten		50 Minuten		50 Minuten	
Netto-Fallzahl	2.036	1.102	2.402	1.116	2.342	1.108
Ausschöpfung	46,9	53,7	54,2	54,2	53,2	55,2

Infratest Sozialforschung 2003: eigene Darstellung

Die für den ALLBUS 2000 aufgeführten Gründe für Nonresponse (Übersicht 2) zeigen, wie bei hart bearbeiteten und dokumentierten Verfahrensmodellen üblich, dass rein quantitativ betrachtet weniger die prinzipielle Nichterreichbarkeit von Zielpersonen, sondern vorrangig eine mangelnde Kooperationsbereitschaft („Verweigerung“) zum Nonresponse führt. Es finden sich keinerlei Hinweise dafür, dass die Stichprobenausfälle durch mangelhaften Aufwand in der Feldbearbeitung zustande gekommen sind oder einem mangelndem Engagement der Interviewerinnen und Interviewer geschuldet wären.¹⁰ Wie bereits

¹⁰ Diese Aussage wird übrigens nicht dadurch relativiert, dass auch bei einer Einwohnermeldeamtsstichprobe ein nicht unerheblicher Teil der Adressen als qualitätsneutrale Ausfälle deklariert

analog bezüglich der generellen Höhe der Ausschöpfungsquote erläutert, lässt auch diese Übersicht noch keine gesicherten Rückschlüsse auf mögliche systematische Effekte zu. Entscheidend ist nicht, welcher der von den Interviewern dokumentierten finalen Ausfallgründe am häufigsten auftritt, sondern inwieweit damit Effekte einhergehen, die zu einer Verzerrung der Stichprobe in einem relevanten Ausmaß führen.

Übersicht 2: Gründe für Nonresponse im ALLBUS 2000

	West	Ost
Bruttostichprobe	100	100
./. stichprobenneutrale Ausfälle		
– Adresse falsch	3,5	2,9
– Zielperson verstorben	0,7	0,9
– Zielperson verzogen	8,0	7,6
– Zielperson lebt nicht mehr in Privathaushalt	1,0	0,9
– Adresse nicht abschließend bearbeitet	1,6	0,2
Bereinigte Bruttostichprobe	100	100
./. Systematische Ausfälle		
– Im Haushalt niemanden angetroffen	5,3	1,4
– Zielperson nicht angetroffen	2,1	1,0
– Zielperson nicht befragungsfähig	1,5	1,8
– Zielperson aus Zeitgründen nicht bereit	3,8	1,6
– Zielperson generell nicht bereit	36,8	37,8
– Zielperson spricht nicht hinreichend gut deutsch	1,5	0,1
– Interview nicht korrekt durchgeführt	2,2	2,6
Auswertbare Interviews	46,9	53,7

Koch/Wasmer/Harkness/Scholz (2001: 58)

Wie ist dieser Gesamtbefund nun zu beurteilen? Führen auch schlechter ausgeschöpfte Zufallsstichproben zu akzeptablen Ergebnissen, oder manövrieren wir uns mit dem „best practice“ Modell für persönlich/mündliche Repräsentativerhebungen immer mehr ins Abseits?

werden. Die vorliegenden Erfahrungen zeigen, dass um die 15 Prozent der Zielpersonen, die anhand der Einwohnermelderegister gezogen wurden, nicht (mehr) unter der angegebenen Adresse wohnen bzw. dort nicht mehr erreichbar sind. Beim ALLBUS-Verfahrensmodell werden diese Adressen, um systematische Effekte zu minimieren, durch Reserveadressen, die bei den gleichen Einwohnermeldeämtern vorab mitgezogen wurden, kontrolliert ersetzt.

Im Folgenden werden wir drei verschiedene Wege zur Beurteilung der Ergebnisqualität von Erhebungen exemplarisch ausführen. In einem ersten Zugang werden im Sinne einer Nettovalidierung einige Randverteilungen der ALLBUS-Erhebungen der Jahre 1994, 1996 und 2000 mit den Ergebnissen des Mikrozensus als externer Referenzstatistik abgeglichen. In einem zweiten Zugang wird danach für den ALLBUS 2000 eine Nonresponse-Analyse durchgeführt. Hierbei wird, neben den verfügbaren Merkmalen, die für die Bruttostichprobe von den Einwohnermeldeämtern geliefert wurden, zusätzlich auf externe (kommerziell vertriebene) mikrogeografische Datensätze, mit denen einige prägende Charakteristika der Straßenzüge, aus denen die Bruttoadressen stammen, beschrieben werden können, sowie auf ausgewählte Merkmale einer Befragung der beim ALLBUS 2000 eingesetzten Interviewerinnen und Interviewer zurück gegriffen. Als drittem Zugang werden abschließend die Effekte der für die ALLBUS-Erhebungen typischen intensiven Nachbearbeitungswellen im Vergleich zu den Ergebnissen der ALLBUS-Hauptbearbeitung analysiert.

Wir knüpfen hierbei auch an die zuletzt ebenfalls von Koch im Jahre 2002 vorgelegte Arbeit „20 Jahre Feldarbeit im ALLBUS: ein Blick in die Blackbox“ an, die zusammen mit den bereits erwähnten Methodenberichten einen umfassenden Einblick in die Realisierungsbedingungen der ALLBUS-Untersuchungen gibt (Koch 2002). Im Unterschied zum letztgenannten Artikel, der die Regelgebundenheit des Verfahrensmodells in den Vordergrund stellt, möchten wir uns vorrangig mit dem Aspekt der Ergebnisqualität befassen.

2. Vergleich der Randverteilungen der ALLBUS-Erhebungen 1994, 1996 und 2000 mit dem Mikrozensus

Im folgenden Abschnitt möchten wir die Ergebnisse der ALLBUS-Erhebungen 1994, 1996 und 2000 im Vergleich zu den Ergebnissen des Mikrozensus des Statistischen Bundesamtes betrachten und bewerten. Hierbei gilt es zu überprüfen, ob sich Belege dafür finden lassen, dass schlechter ausgeschöpfte Erhebungen im Vergleich zum Mikrozensus größere Abweichungen aufweisen. Dieses mit dem Begriff der Nettovalidierung zur charakterisierende Verfahren entspricht in etwa dem Vorgehen in der bereits erwähnten Arbeit von Koch (1998), der auf diese Weise verschiedene Zufallsstichproben, für die unterschiedliche Ausschöpfungsquoten berichtet wurden, untersucht hat.¹¹

11 Vergleiche zum Verfahren und zu den Ergebnissen Koch (1998) sowie vorher Hartmann (1990) oder Hartmann/Schimpl-Neimanns (1992).

Wie bereits ausgeführt, ist die Ausschöpfung bei den letzten drei ALLBUS-Erhebungen, die anhand von Einwohnermeldeamtsstichproben gebildet wurden, weiter gesunken. Vor allem in den alten Bundesländern hat sich seit 1994 eine Verringerung der Ausschöpfung um mehr als 6 Prozentpunkte ergeben. Mit 1,5 Prozentpunkten fällt der Rückgang in den neuen Bundesländern weniger dramatisch aus. Um die möglichen Auswirkungen besser unterscheiden zu können, haben wir den Abgleich mit der amtlichen Statistik immer getrennt nach Ost und West vorgenommen. Die untersuchten Merkmale umfassen Alter, Geschlecht, Familienstand, Erwerbsstatus, Schulabschluss und Haushaltsgröße. Weitere Validierungen sind nicht möglich, da der Mikrozensus ansonsten über keine mit dem ALLBUS direkt vergleichbaren Merkmale verfügt. Die Ergebnisse der Validierung haben wir in der folgenden Übersicht 3 zusammengefasst. Zur Veranschaulichung haben wir dabei für jede der abzugleichenden Merkmalsausprägungen alle positiven Abweichungen in den ALLBUS-Untersuchungen, also alle Ausprägungen, die größer als die Referenzergebnisse im Mikrozensus ausfallen, aufaddiert und in einem dadurch komprimierten Gesamtüberblick dargestellt. Die tabellarischen Nachweise mit den Einzelergebnissen sind im Anhang beigefügt.

Insgesamt weisen die ALLBUS-Stichproben eine recht gute Anpassung an die Verteilungen der amtlichen Statistik auf. In den demografischen Grundverteilungen (Alter, Familienstand, Geschlecht sowie Ein- bzw. Mehrpersonenhaushalte) schwanken die Abweichungen in der Regel mit nicht mehr als 2 bis 3 Prozentpunkten um das Referenzergebnis.

Beim höchsten Schulabschluss und beim Erwerbsstatus ergeben sich zum Teil etwas größere Abweichungen, was jedoch auch methodisch verursacht und mit abfragebedingten Artefakten zusammenhängen könnte. Der Mikrozensus weist beim Schulabschluss, insbesondere bei älteren Personen, zu einem nicht unerheblichen Anteil fehlende Angaben auf, die die Gesamtverteilung verzerren. Beim Erwerbsstatus ist davon auszugehen, dass das Berichtswochenkonzept des Mikrozensus, bei der eine konkrete Erwerbstätigkeit in einer definierten Woche abgefragt wird, im Vergleich zum ALLBUS zu systematisch höheren und damit genaueren Angaben führt, da hierbei Teilzeit-Beschäftigung sowie insbesondere geringfügige Beschäftigungsverhältnisse präziser abgebildet werden. Entsprechende methodische Effekte können sich leicht zu Abweichungen in Höhe von mehreren Prozentpunkten addieren.¹²

12 Vergleiche dazu auch die methodischen Anmerkungen von Koch (1998) in den Fußnoten 18 bis 20.

Übersicht 3: Vergleich der Summe der Abweichungen aus ALLBUS und Mikrozensus (Addition der positiven Abweichungen in Prozentpunkten)

	ALLBUS 2000	ALLBUS 96	ALLBUS 94
West			
Ausschöpfungsquote	46,9	54,2	53,2
Geschlecht	0,3	2,2	3,8
Alter	2,6	3,3	2,9
Familienstand	0,6	2,8	1,6
Erwerbsstatus	2,5	1,7	4,4 ¹⁾
Schulabschluss	4,1	7,7	5,7 ¹⁾
Haushaltsgröße	3,0	3,9	1,5
Ost			
Ausschöpfungsquote	53,7	54,2	55,2
Geschlecht	0,1	0,2	0,6
Alter	2,5	3,5	4,5
Familienstand	0,9	5,4	3,7
Erwerbsstatus	1,0	0,7	0,2 ¹⁾
Schulabschluss	3,8	2,4	4,3 ¹⁾
Haushaltsgröße	0,1	3,0	2,9

1) Referenz: Mikrozensus 93

Infratest Sozialforschung, eigene Berechnungen

Insgesamt lässt sich feststellen, dass die sinkende Ausschöpfung bei den letzten ALLBUS-Erhebungen nicht mit einer Zunahme der Abweichungen bei den soziodemographischen Randverteilungen einhergeht. Im Gegenteil! Der ALLBUS 2000 weist sowohl in den alten als auch in den neuen Bundesländern bei den soziodemografischen Grundmerkmalen insgesamt die geringsten Abweichungen auf.

Eine geringfügig andere Tendenz findet sich beim Erwerbsstatus. Hier weicht der ALLBUS 2000 etwas weiter vom Mikrozensus ab, als dies im Westen 1996 bzw. im Osten bei beiden vorherigen Erhebungen der Fall gewesen ist. Im Einzelnen wurden im Westen im Jahr 2000 2,5 Prozentpunkte mehr sog. Nichterwerbspersonen und im Osten 1 Prozentpunkt mehr Arbeitslose als im Mikrozensus ausgewiesen (vergleiche dazu die tabellarische Übersicht im Anhang). Die Abweichungen sind für sich genommen allerdings nicht besonders hoch ausgeprägt.

Weitere kleinere Abweichungen beziehen sich auf das Merkmal Haushaltsgröße, bei dem im ALLBUS 2000 in den alten Bundesländern die Abweichung gegenüber dem ALLBUS 1996 geringer, gegenüber dem ALLBUS 1994 jedoch höher ausfallen. Dies lässt sich auch

beim Merkmal Schulabschluss in den neuen Bundesländern feststellen, wo die Abweichung vom ALLBUS 2000 gegenüber dem ALLBUS 1994 geringer, gegenüber dem ALLBUS 1996 jedoch höher ausfällt. Dahingegen gilt, dass in den alten Bundesländern bei den Merkmalen Alter, Geschlecht, Familienstand und Schulabschluss die Abweichungen beim ALLBUS 2000 geringer als in den vorherigen Studien waren. In den neuen Bundesländern lässt sich dies beim ALLBUS 2000 für die Merkmale Alter, Geschlecht, Familienstand und Haushaltsgröße feststellen.

Anhand des Vergleichs der ALLBUS-Erhebungen 1994, 1996 und 2000 mit der amtlichen Statistik lassen sich demnach keine hinreichend eindeutigen empirischen Belege für die Hypothese finden, dass schlechter ausgeschöpfte Stichproben generell größere Abweichungen nach sich ziehen. Vielmehr zeigt sich, dass der ALLBUS 2000 trotz der relativ geringen Ausschöpfung in sehr weiten Teilen weniger Abweichung zur amtlichen Statistik aufweist als die vorherigen ALLBUS-Untersuchungen der Jahre 1996 und 1994. Auch die Unterscheidung zwischen den alten und neuen Bundesländern hat nicht bestätigen können, dass mit einer besseren Ausschöpfung in den neuen Bundesländern eine bessere Annäherung an die amtliche Statistik verbunden wäre.

Die nach einem hart dokumentierten und registerbasierten Verfahrensmodell durchgeführten ALLBUS-Untersuchungen der Jahre 1996 und 2000 haben demnach den mit dem ALLBUS 1994 im Vergleich zu anderen, besser ausgeschöpften allgemeinen Bevölkerungsumfragen gesetzten Standard bezüglich der Übereinstimmung der soziodemografischen Grundverteilungen mit der amtlichen Statistik nicht nur gehalten, sondern sogar noch leicht verbessert.

3. Effekte von Nonresponse beim ALLBUS 2000

Die Analyse der Randverteilungen bei den ALLBUS-Untersuchungen ermöglicht eine grundsätzliche Verortung der Stichprobe anhand einiger zentraler soziodemografischer Grundparameter. Dies ist insbesondere dann bedeutsam, wenn die Ergebnisse auf die Bevölkerung insgesamt hochgerechnet werden sollen und Quantifizierungen vorgesehen sind. Auf der anderen Seite lässt sich aus einer akzeptablen Abbildung der Gesamtstruktur der Bevölkerung noch nicht hinreichend genau ableiten, dass deshalb auch andere, zum Beispiel komplexere Einstellungsmerkmale in der Stichprobe richtig verteilt sein müssen. Hierfür ist es nötig, zu überprüfen, ob es gelungen ist, im Rahmen der Feldarbeit auch solche Bevölkerungsgruppen zu erreichen, die in einem relevanten Anteil zur Grundge-

samtheit gehören,¹³ dabei jedoch aufgrund bestimmter inhaltlich begründbarer Faktoren schwerer zu befragen sind.

Derartige Fragestellungen lassen sich – mangels verfügbarer externer amtlicher Statistiken – nur auf Basis einer Nonresponse Analyse genauer beurteilen. Ihr Gegenstand besteht, vereinfacht ausgedrückt darin, zu prüfen, inwieweit sich die Verteilung der Respondenten von der Verteilung der Nonrespondenten bezüglich bestimmter vorab festgelegter Merkmale signifikant und bedeutsam unterscheidet, die als typisch für Nonrespondenten angesehen werden und die damit zu einem systematischen "nonresponse-bias" führen können.¹⁴

3.1. Typisierung von Nonresponse

Aus unserer Sicht lassen sich – in Anlehnung an Schnell (1997) – folgende sechs empirisch operationalisierbare Problembereiche, die zu einem stärkeren Nonresponse bei allgemeinen Bevölkerungsumfragen in Privathaushalten führen können, abgrenzen.

- **Unterschiedliche Erreichbarkeit**

Grundsätzlich kann davon ausgegangen werden, dass sich Personen in Privathaushalten je nach Statuspassage („live cycle factors“ – Westerhoven 1978) bzw. Lebensstil hinsichtlich ihrer Antreffbarkeit zu Hause unterscheiden. Typische empirische Indikatoren für „schlechtere“ Antreffbarkeit sind im Rahmen einer Nonresponse Analyse zum Beispiel die Haushaltsgröße (Single-Haushalte) und/oder der Erwerbsstatus (Vollzeit-Erwerbstätig bzw. Studium) bzw. – in Korrelation – die Zugehörigkeit zu bestimmten Altersgruppen, insbesondere jüngere Menschen zwischen Anfang 20 und Mitte 30 Jahren.

13 Diese Bedingung schließt zum Beispiel „Obdachlose“, Personen in Justizvollzugsanstalten, Bewohnerinnen und Bewohner von Pflegeheimen etc., bzw. bei Einwohnermeldeamts basierten Stichproben auch generell Personen, die in Deutschland nicht gemeldet sind, aus. Sie gehören definitionsgemäß nicht zur Grundgesamtheit der „Wohnbevölkerung in Privathaushalten“. Zum „Coverage“-Problem vergleiche Schnell (1991).

14 Als die nach wie vor umfassendste Studie zum Problem des Nonresponse kann für den deutschsprachigen Raum die von Schnell (1997) betrachtet werden. Die von uns gewählte Typisierung der Problembereiche, die mit einem stärkeren Nonresponse verbunden sein können, beruht im Wesentlichen auf den Ausführungen in dieser Arbeit. Ein Überblick findet sich ebenfalls bei Groves (1989: 186ff).

- **Krankheit und Gebrechlichkeit**

Abgesehen von dem Sonderproblem der Nicht-Befragbarkeit von bestimmten Bevölkerungsgruppen in persönlich-mündlichen Erhebungen (Taubstumme, schwerhörige- bzw. sprachbehinderte Personen, geistig Behinderte, dementiell Erkrankte etc.) finden sich bei älteren bzw. hochbetagten Menschen deutlich häufiger chronische Erkrankungen, die zu einem Nonresponse führen können. Relevanter empirischer Indikator hierfür ist das Alter, sprich die Zugehörigkeit zur Altersgruppe ab 65 bzw. ab 70 Jahren.

- **Fehlende Kenntnisse der deutschen Sprache**

Fehlende Sprachkenntnisse sind insbesondere für ausländische Mitbürgerinnen und Mitbürger sowie für Zuwanderer eine Hürde, die zu einem verstärkten Nonresponse führen kann. Indikatoren hierfür sind die Nationalität sowie das Herkunftsland.¹⁵

- **Misstrauen aufgrund von Viktimisierungsängsten**

Kriminalitätsfurcht sowie daraus resultierendes Misstrauen und fehlende „Hilfsbereitschaft“ gelten insbesondere in Ballungsräumen bzw. städtischen Kerngebieten als weiterer Erklärungsansatz für Nonresponse. Indikatoren hierfür sind der Siedlungsstrukturtyp bzw. die Wohnquartiersform (Hochhäuser).

- **Umfragen als „Belastung“ bzw. kritische Distanz gegenüber der Datenerhebungen (Schutz der Privatsphäre, Datenschutz, „politisch“ motiviertes Misstrauen)**

Nonresponse als Ausdruck einer Belastung ist dann gegeben, wenn aus der Sicht des Befragten die Teilnahme im Widerspruch zu wichtigen persönlichen Zielen steht oder dadurch Nachteile erwartet werden. Dies ist zum Beispiel dann gegeben, wenn die Zielperson vermutet, dass die Anonymität nicht gewährleistet ist. Entsprechend motiviertes Misstrauen, das sich z. T. natürlich auch auf reale Erfahrungen (als Umfrage getarnte „kalte“ Akquise, Weiterverkauf von Adressen zum Zwecke des gezielten Marketings etc.) stützen kann, wird insbesondere bei bestimmten Bevölkerungsgruppen aus dem intellektuellen bzw. alternativen Milieu vermutet, denen vor dem Hintergrund der Erfahrungen mit der Volkszählung von 1987 nach wie vor eine besondere Distanz gegenüber Datenerhebungen nachgesagt wird. Indikatoren hierfür sind zum Beispiel Milieuvariablen oder postmaterialistische Präferenzen.

15 Wie bereits erwähnt, gehören nicht in Privathaushalten gemeldete Personen nicht zur Grundgesamtheit von einwohnermeldeamtsbasierten Studien. Das Problem der fehlenden Sprachkenntnisse von zum Beispiel illegal eingewanderten Personen oder von Asylanten entfällt damit an dieser Stelle.

- **Fehlendes Interesse bzw. soziale Distanz**

Eine subjektiv geringere Bedeutsamkeit des Gegenstandes der Untersuchung bzw. ein postuliertes fehlendes Interesse an sozialwissenschaftlichen Themenstellungen aufgrund von Deprivation, fehlender Bildung oder eigener sozialer Distanz stellt eine besondere Variante eines für die professionelle empirische Sozialforschung (angeblich) typischen „Middle-Class-Bias“ dar (Esser/Grohmann/Müller/Schäffer 1989). Typische Indikatoren hierfür sind Bildung, Schichtzugehörigkeit oder auch Interesse an Politik und/oder Gesellschaft.

Im Rahmen einer empirischen Analyse ist demnach zu prüfen, inwieweit sich Respondenten und Nonrespondenten signifikant und bedeutsam in einem oder mehreren der genannten Merkmalsbereiche unterscheiden. Da sich die genannten Bereiche zum Teil überlappen, in Wechselwirkung zueinander stehen oder sich auch gegenseitig aufheben können (Beispiel: bessere Erreichbarkeit von Älteren versus chronische Krankheiten), ist es geboten, dies möglichst im Rahmen einer multivariaten statistischen Analyse zu tun.

3.2 Verfügbare Merkmale

Kernproblem jedweder Nonresponse Analyse ist die praktisch verursachte Schwierigkeit, in der Regel nur für Respondenten, jedoch nicht für Nonrespondenten über geeignete empirische Merkmalsausprägungen zu verfügen.¹⁶ Wir haben in unserer Arbeit einen besonderen (und sicherlich methodisch auch nicht gänzlich unproblematischen) Weg beschritten, um dieses grundsätzliche Problem mit einem alternativen Design zu umgehen.

In Ergänzung zu den im Rahmen der Ziehung der Adressen von den Einwohnermeldeämtern mitgelieferten Merkmale Alter, Geschlecht und Nationalität sowie dem regionalen Schichtungsmerkmal BIK-Siedlungsstrukturtyp (Kerngebiet, Randlage von Stadtregionen, Peripherie, Ländlicher Raum) haben wir die Bruttoadressen zusätzlich mit den extern verfügbaren und kommerziell vertriebenen mikrogeografischen Strukturmerkmalen der Fa. Microm – den so genannten MOSAIC Informationen – angereichert.¹⁷ Hierbei han-

16 Völlig zu Recht stellt auch Schnell fest, dass die (...) „empirische Analyse von individuellem Nonresponse (...) vor dem Problem (steht), dass Nonrespondenten gerade darüber definiert werden, dass für Sie weniger (bzw. i.d.R. sogar überhaupt keine – US/IL) Daten vorliegen, als für den Rest der untersuchten Population.“ (Schnell 1997: 134).

17 Microm GmbH und Sinus Sociovision GmbH: MOSAIC Informationen. Stand 2000. Neuss, Frühjahr 2000. Seit dem Jahr 2002 liegt eine überarbeitete Version vor, mit der die neue Typologie von Sinus Sociovision nun auch für die neuen Bundesländer implementiert ist. Wir beziehen uns noch auf die alte Typologie, da wir die Informationen bereits im Jahre 2001 den

delt es sich um eine Datenbank, die das Wohnumfeld der einzelnen Haushalte in Deutschland beschreibt. In der Basisdatenbank sind Informationen zu folgenden Sachverhalten enthalten (MOSAIC Variablen):

- Art der Bebauung und Quartierstypen,
- Sozioökonomische Merkmale (Status, Ausländeranteil, Anonymitätsbedürfnis, Kreditausfallrisiko),
- Demografie (Alter, Familienstruktur).

Die Informationen sind mit Hilfe eines methodisch leider nicht näher erläuterten mikrogeografischen Segmentierungsverfahrens so verdichtet worden, dass sie als Umfeldcharakteristik prinzipiell jedem Haushalt in Deutschland zugespielt werden können.

Darüber hinaus stehen verschiedene Typologien zur Verfügung. Interessant für unsere Analyse sind die so genannten Sinus-Milieus. Hierbei werden Personengruppen anhand von typischen Einstellungen und Wertorientierungen zu 10 Milieus zusammengefasst, die sich nach der jeweiligen sozialen Lage bzw. Schichtzugehörigkeit (Unterschicht bis Oberschicht) sowie nach ihrer Grundorientierung von „traditionell“ bis „postmodern“ unterscheiden (vergleiche dazu die Kurzcharakterisierung im Anhang). Die Sinus-Typologie ist im Bereich der sozialwissenschaftlichen Forschung bisher eher unüblich und von daher wenig rezipiert. Mit offenbar großem Erfolg wird die Typologie hingegen seit Jahren insbesondere im Bereich der Medienforschung eingesetzt.¹⁸ Der Informationsbestand der Microm-Datenbank wurde mit den Sinus-Milieus zu den MOSAIC Milieus verknüpft. Dazu wurden ebenfalls nicht näher spezifizierte Modelle entwickelt, welche für jedes Haus die statistische Wahrscheinlichkeit berechnen, mit der die Milieus darin vorkommen, und welches davon das dominante Milieu ist. Bis 2002 waren die Sinus-Milieus in Form der MOSAIC Adaption nur für Westdeutschland verfügbar. Aus diesem Grunde können wir die Sinus-Milieus im Rahmen der Nonresponse Analyse auch nur für die alten Bundesländer (inkl. Berlin/West) ausweisen.¹⁹

Natürlich müssen die angereicherten Daten der Fa. Microm vorsichtig beurteilt werden. Sie bilden das Wohnumfeld der Haushalte ab und basieren selber auf diversen aggregier-

ALLBUS-Daten angereichert haben. Vergleiche dazu auch die Informationen auf der Website der Fa. MICROM unter <http://www.microm-online.de>.

18 Bei den Sinus-Milieus handelt es sich um ein urheberrechtlich geschütztes Verfahren der Fa. Sinus Sociovision GmbH. Eine Referenz- und Literaturliste findet sich auf der Homepage von Sinus Sociovision unter <http://www.sinus-milieus.de>.

19 Eine nachträgliche Anreicherung der ALLBUS-Daten mit den neuen MOSAIC Milieus ist nicht möglich, da die Adressen des ALLBUS 2000 aus datenschutzrechtlichen Gründen bereits vernichtet wurden.

ten Datenbeständen, die zu diesem Zwecke zusammengeführt wurden. Wir möchten an dieser Stelle jedoch noch einmal den eigentlichen Zweck dieses Zugangs in Erinnerung rufen. Es geht uns darum, relevante und im ALLBUS ansonsten nicht vorhandene Variablen zu generieren, mit denen sich Respondenten und Nonrespondenten hinsichtlich bestimmter Merkmale beschreiben und ggf. voneinander abgrenzen lassen. Darüber hinaus haben wir dem ALLBUS-Datensatz noch zusätzlich ausgewählte Merkmale einer von uns im Frühjahr 2001 durchgeführten Interviewerbefragung beigelegt.²⁰ Die ausgewählten Variablen bilden näherungsweise die in der Literatur häufig als relevant betrachteten interviewerbezogenen Einflussfaktoren „Persönliche Interviewermerkmale“, „Erfahrung“, „Art der Kontaktstrategie“ sowie „Organisation des Feldeinsatzes“ ab.²¹ Übersicht 4 gibt einen Gesamtüberblick zu den im Rahmen der Nonresponse Analyse verfügbaren Variablen.

Da sich im Rahmen der Analyse keinerlei Hinweise darauf fanden, dass die zugespielten Wohnumfeldinformationen nicht trennscharf sind, spricht aus unserer Sicht nichts dagegen, in diesem Fall diesen ansonsten eher unüblichen und methodisch noch nicht validierten Zugang der Anreicherung einer Stichprobe mit externen Datenbeständen zu wählen.²²

20 Befragt wurden 451 ALLBUS-Interviewer. Mit 392 auswertbaren Fragebögen betrug die Rücklaufquote 87 Prozent.

21 Vergleiche hierzu im Überblick ebenfalls Schnell (1987: 195ff).

22 Eine Korrelationsanalyse der angereicherten MOSAIC Daten mit den Ergebnissen des ALLBUS (Netto-Stichprobe) ergab hoch korrelierte Zusammenhänge zum Beispiel zwischen MOSAIC Lebensform (Single- versus Familiendichte) und den ALLBUS-Daten zur Haushaltsgröße, zum Familienstand und zur Zahl der Kinder im Haushalt. Ebenfalls gut korreliert ist das Merkmal MOSAIC Bebauungstyp mit der ALLBUS-Variable Art des Gebäudes gemäß vorliegender Interviewereinstufung. Da die Interviewereinstufung zur Art des Gebäudes im ALLBUS 2000 für einen Großteil (knapp 40 Prozent) der Nonrespondenten leider nicht vorliegt, konnte auf diese und einige weitere unmittelbar erhobene Wohnquartiersmerkmale nicht zurückgegriffen werden.

Übersicht 4 : Response versus Nonresponse – Verfügbare Merkmale und Operationalisierung

Testhypothese:

Respondenten und Nonrespondenten unterscheiden sich signifikant und hinreichend stark bezüglich einer oder mehrerer der folgenden Dimensionen.

Sample-Point bezogene Faktoren

1. *Unterschiedliche Erreichbarkeit*
 - MOSAIC Lebensform **Single-Dichte** versus Familien-Dichte im Wohngebiet
 - MOSAIC Bebauungstyp **Wohnblocks/Hochhäuser** versus mittelgroße- / Familienwohnhäuser
 - BIK-Typ **Kerngebiet** versus Randlage / Peripherie / ländlicher Raum
2. *Gebrechlichkeit / Krankheit*
 - Alter **Zielpersonen ab 70 Jahre** versus sonstige Altersgruppen
3. *Fehlende Kenntnisse der deutschen Sprache*
 - Nationalität **Nicht-Deutsch** versus Deutsch
4. *Misstrauen aufgrund von Viktimisierungserfahrungen*
 - BIK-Typ **Kerngebiet** versus Randlage / Peripherie / ländlicher Raum
 - MOSAIC Bebauungstyp **Wohnblocks/Hochhäuser** versus mittelgroße bzw. Familienhäuser
5. *Belastung bzw. kritische Schutz der Privatsphäre*
 - Prägendes Sinus-Milieu **Intellektuelles Milieu** versus statusorientierte- / adaptive- / sonstige Milieus (nur für die alten Bundesländer verfügbar)
 - Anonymitätsbedürfnis **Wohngebiete mit überdurchschnittlichem Anonymitätsbedürfnis** versus unterdurchschnittlicher Anonymität
6. *Fehlendes Interesse und/oder generelle soziale Distanz*
 - MOSAIC Status **Wohngebiete mit unterdurchschnittlichem Status** versus Gebiete mit überdurchschnittlichem Status
 - MOSAIC Bebauungstyp **Wohnblocks/Hochhäuser** versus mittelgroße bzw. Familienhäuser

Interviewerbezogene Faktoren

1. *Persönliche Merkmale*
 - Intervieweralter **65 u. älter** versus Mittelalt / Jung
 - Schulabschluss **Hochschule/Abitur** versus mittlere Reife / Hauptschule
 - Geschlechterkongruenz **Gleich-** versus gegengeschlechtliche Interviewsituationen
2. *Erfahrung / Kontaktverhalten*
 - Interviewererfahrung **10 u.m. Jahre** versus kürzere Erfahrungsjahre
 - Präferiertes Verfahren **AR** versus RR / Quote
 - Art der Kontaktaufnahme **Interview ohne Vorankündigung: oft** versus gelegentlich/selten
3. *Feldorganisation*
 - Einsatzentfernung **unter 40 km** versus 40 km und mehr vom eigenen Wohnort entfernt
 - Arbeitsbelastung **Zeit bis zum ersten realisierten Interview im Point:**
max. 2 Wochen versus längerer Bearbeitungsdauer

4. Ergebnisse

Auf Basis der vorhandenen Variablen haben wir 3 Modelle berechnet, mit denen der Einfluss der Merkmale auf die Verteilung der Bruttostichprobe nach Response und Nonresponse geschätzt werden kann. Das von uns gewählte Design einer multivariaten logistischen Regression (Logit-Modell) ermöglicht es, zu überprüfen, inwieweit der Stichproben-Nonresponse beim ALLBUS 2000 auf eine oder mehrere der vorab genannten Stichprobenausfälle zurückzuführen ist.²³ Ergibt sich ein statistisch signifikanter und hinreichend starker Zusammenhang, so muss von einer systematischen Verzerrung, sprich von einem „non response bias“ ausgegangen werden. Lässt sich dies nicht nachweisen, so kann dies aus unserer Sicht als Beleg dafür gewertet werden, dass sich der Stichproben-Nonresponse weitgehend unsystematisch und damit gemäß der Grundlogik einer zufallsbasierten Stichprobe verteilt. Die Ergebnisse der Nonresponse-Analyse sind in der folgenden Übersicht 5 zusammengefasst. Ausgewiesen sind die Regressionskoeffizienten sowie das statistische Signifikanzniveau.²⁴ Als hinreichend gesichert sind grundsätzlich nur statistisch signifikante Ausprägungen zu betrachten. Das Vorzeichen der Regressionskoeffizienten gibt an, inwieweit die Merkmalsausprägung die Wahrscheinlichkeit in Richtung Nonresponse (-) oder Response (ohne Vorzeichen) verändert. Da es sich bei der Mehrzahl der Variablen um kategoriale Merkmale handelt, ist ggf. die Referenzkategorie, auf die sich die Effekte der anderen Ausprägungen im Vergleich beziehen (Dummy bzw. Indikator-Codierung), gleich zu Beginn benannt. Zur Beurteilung der Schätzstärke, sprich der Erklärungskraft des Gesamtmodells, ist am Ende das sog. Pseudo- R^2 (nach McFadden) ausgewiesen.²⁵ Laut Andreß/Hagenaars/Kühnel deuten Werte unter 0,05 auf einen eher geringen und demnach zu vernachlässigenden Zusammenhang sowie Werte über 0,2 auf einen starken Zusammenhang. Werte über 0,4 sind hingegen nur sehr selten zu erreichen (Andreß/Hagenaars/Kühnel 1997: 288).

23 Mit Hilfe eines multivariaten Logit-Modells lässt sich prüfen, inwieweit bestimmte Variablen, die eines oder mehrere der vorab spezifizierten Konstrukte operationalisieren, als stabile Prädiktoren für Nonresponse zu interpretieren sind. Sofern die Variablenausprägungen signifikant und das Modell auch insgesamt trennscharf genug ist (goodness of fit), kann dies im Unterschied zu einfachen bivariaten signifikanten Abweichungen als systematische Determination interpretiert werden. Zur Notwendigkeit von multivariaten Analysen vergleiche die Arbeit von Rendtel (1995: 198ff).

24 Aus Gründen der Übersichtlichkeit haben wir auf die Ausweisung von Logits verzichtet. Alle Modelle wurden mit der Prozedur NomReg unter SPSS 11.0 berechnet.

25 Das Pseudo- R^2 gibt an, um wie viel sich der Fehler des Modells unter dem Einfluss der Prädiktoren reduziert (PRE: proportional reduction of error). Es lässt sich demnach analog zum Determinationskoeffizienten der linearen Regression als Grad der Varianzaufklärung interpretieren (Andreß/Hagenaars/Kühnel 1997: 287f).

Übersicht 5: Nonresponse Analyse: ALLBUS 2000, alte Bundesländer

Logistische Regression: Nonresponse versus Response

	Modell 1 Einfache Merkmale der Melderegister		Modell 2 Erweiterte Merkmale inkl. MOSAIC Daten		Modell 3 Sample-Point- und Interviewermerkmale	
	West	Ost	West	Ost	West	Ost
Geschlecht	,082	,060	-,068	,068	(redund.)	(redund.)
Altersklassen Ref.-Kat.: 70 und älter						
18-29	,453**	,515*	,432**	,500**	,527**	,749**
30-39	,148	,317	,125	,294	,238	,505*
40-49	,269*	,103	,246*	,056	,374**	,144
50-59	,253*	,283	,211	,251	,312*	,422*
60-69	-,001	,237	-,021	,202	,040	,106
Keine Angabe (EWMA)	,572**	-,053	,508*	,021	,170	-,060
Nationalität Ref.-Kat.: Nicht Deutsch						
Deutsch	-,066	-,048	-,130	,016	-,227	-,028
Keine Angabe (EWMA)	,128	,229	,082	,332	,006	,335
BIK-Typ Ref.-Kat.: Kerngebiet						
Randlage Stadtregion	,218*	,350*	,097	,310	-,000	,872**
Peripherie	-,195*	,026	-,336**	-,023	-,347*	,028
Ländlicher Raum	,228**	,643**	,045	,645**	-,109	,816**
MOSAIC Lebensform Ref.-Kat.: überdurchschn. Singles						
leicht überdurchschn. Singles			,064	,457*	,069	,560*
gemischtes Wohngebiet			,039	,295	-,007	,248
leicht überdurchschn. Familien			,064	,345	-,052	,145
überdurchschn. Familien			,093	,001	,114	-,023
Keine Angabe			-,253	(redund.)	,113	(redund.)
Mosaic Bebauungstyp Ref.-Kat.: Wohnblöcke						
1-2 Familienhäuser			,233	,262	,348*	,412*
3-5 Familienhäuser			,036	-,010	,139	,206
6-9 Familienhäuser			-,024	,251	,200	,482*
Keine Angabe			-,150	-,055	-,453	-,421
MOSAIC Status Ref.-Kat.: unterdurchschn. Status						
leicht unterdurchschn. Status			,038	,210	,086	,216
durchschnittl. Status			,044	,100	-,039	,162
leicht überdurchschn. Status			,132	,133	,075	,218
überdurchschn. Status			,059	,338*	-,041	,373
Keine Angabe			(redund.)	(redund.)	(redund.)	(redund.)

Fortsetzung Nonresponse Analyse:						
	Modell 1		Modell 2		Modell 3	
	Einfache Merkmale der Melderegister		Erweiterte Merkmale inkl. MOSAIC Daten		Sample-Point- und Interviewermerkmale	
	West	Ost	West	Ost	West	Ost
Prägendes Sinus-Milieu (West)						
Ref.-Kat.: Intellekt. Milieu						
Etabliertes Milieu			,001	-	,118	-
Postmodernes Milieu			-,096	-	-,178	-
Modernes bürgerl. Milieu			,031	-	,219	-
Adaptives Milieu			,069	-	-,070	-
Statusorientiertes Milieu			,325*	-	,280	-
Trad. bürg. Milieu			,044	-	,122	-
Trad. Arbeitermilieu			-,232	-	-,228	-
Hedonistisches Milieu			,203	-	,236	-
Konsum-materialist. Milieu			,048	-	-,053	-
Keine Angabe			1,471	-	1,611	-
Anonymitätsbedürfnis						
Ref.-Kat.: überdurchschnittl.						
Unterdurchschnittlich			-,054	,150	,136	,134
Leicht unterdurchschnittlich			-,131	,118	,024	,153
Mittlere Anonymität			,080	-,033	,226	-,012
Leicht überdurchschnittl.			-,138	,009	,006	,000
Keine Angabe			-,749	(redund.)	-,618	(redund.)
Alter des Interviewers						
Ref.-Kat.: 65 und älter						
-39					-,231	-,899*
40-54					,119	,220
55-64					,195	-,141
Schulabschluss Interviewer						
Ref.-Kat.: Abitur						
Volksschule					,172	,311
Mittlere Reife					-,014	-,085
Geschlechtsskongruenz						
Ref.-Kat.:						
Z-weiblich / I-weiblich						
Z-männlich / I-männlich					-,046	-,123
Z-männlich / I-weiblich					-,120	,095
Z-weiblich / I-männlich					-,220*	-,049
Interviewererfahrung						
Ref.-Kat.: 10 u.m. Jahre						
bis 2 Jahre					-,254	,636
3-5 Jahre					-,089	-,259
6-9 Jahre					,101	-,191
Präferiertes Verfahren						
Ref.-Kat.: Feste Adressen						
Random-Route					-,106	-,477**

sind im Vergleich zu 18-29jährigen und in den alten Bundesländern auch zu den 40-49jährigen signifikant weniger bei der Gruppe der Respondenten vertreten. Dieser Effekt war für ältere Zielpersonen zu erwarten und spricht, da der ALLBUS als allgemeine Bevölkerungsumfrage nicht gesondert auf Erhebung in Haushalten von hochbetagten Menschen optimiert ist, nicht grundsätzlich gegen die Qualität der Umfrage. Demgegenüber weist die Variable *BIK* nur in den neuen Ländern auf einen etwas stärkeren Nonresponse in Kerngebieten im Vergleich zum ländlichen Raum hin. Andere Indikatoren, wie zum Beispiel der *MOSAIC* *Bebauungstyp* sind hingegen zum Beispiel bezüglich Ausschöpfung in Hochhausgebieten nicht signifikant. Bei den *Sinus-Milieus*, die zum Zeitpunkt der Anreicherung der Daten nur in den alten Ländern verfügbar waren, findet sich beim intellektuellen Milieu nur im Vergleich zum statusorientierten Milieu ein etwas stärkerer Nonresponse. Dieser eher singuläre Effekt lässt sich jedoch kaum als Beleg für eine mögliche Unterausschöpfung von „befragungskritischen“ Zielpersonen aus postmaterialistischen Milieus interpretieren. Gleiches gilt für die zwei weiteren signifikanten Effekte in den neuen Ländern. Die leicht stärkere Ausschöpfung bei Zielpersonen mit *überdurchschnittlichem Status* ist für sich genommen genauso singulär wie die etwas bessere Ausschöpfung von Zielpersonen aus *leicht überdurchschnittlichen Single-Wohngebieten* im Vergleich zu *überdurchschnittlichen Single-Wohngebieten*.

Im **dritten Modell** werden zusätzlich einige relevante Interviewermerkmale mit einbezogen.²⁶ Wir wechseln damit in der Betrachtung die Perspektive. Hatten vorher die Effekte von Nonresponse im Vordergrund gestanden, die zu systematischen Verzerrungen führen können, so geht es jetzt primär um die unterschiedlichen Einflussfaktoren – Sample-Point bezogene und / oder interviewerbezogene – die den beim ALLBUS 2000 festgestellten hohen Nonresponse mit beeinflusst haben.²⁷ Für die alten Bundesländer ergibt sich auch

26 Methodisch muss an dieser Stelle angemerkt werden, dass sich Modell 3 im Vergleich zu den beiden anderen Logit-Modellen auf eine geringere Fallzahl von $n = 3.289$ im Vergleich zu $n = 4.400$ für die alten Länder bzw. $n = 1.529$ im Vergleich zu $n = 2.050$ für die neuen Länder stützt. Die Reduktion der Fallzahl ergibt sich aufgrund der Zuspierung der Ergebnisse der Interviewerbefragung, bei der nicht für alle ALLBUS-Interviews gültigen Angaben realisierbar waren. Streng genommen können deshalb auch die Ergebnisse des dritten Modells aufgrund seiner reduzierten Fallzahl nicht unmittelbar mit den beiden anderen Modellen verglichen werden. Da sich jedoch auch dann, wenn man alle drei Modelle mit der gleichen reduzierten Fallzahl berechnet, im Ergebnis keine relevanten Unterschiede ergeben, haben wir aus Gründen der Übersichtlichkeit trotzdem diese Form der vergleichenden Darstellung inklusive der Ergebnisse des dritten Modells gewählt.

27 Die Verteilung der Interviewer auf die einzelnen Sample-Points erfolgt bei allgemeinen Bevölkerungsumfragen nicht systematisch. Eingesetzt werden im Prinzip die CAPI-Interviewer, die vor Ort verfügbar sind. Im Prinzip kann die Verteilung der Interviewer auf die Sample-Points von daher als im weitesten Sinne randomisiert betrachtet werden. Abgewichen wurde hiervon erst im

bei dieser erweiterten Betrachtung mit einem Pseudo- R^2 von knapp unter 0,05 ein nach wie vor schwacher und damit wenig aussagekräftiger Zusammenhang. Für die neuen Bundesländer wächst das Pseudo- R^2 hingegen auf einen akzeptableren Wert von knapp unter 0,1.

Abgesehen von der bereits erwähnten Unterrepräsentierung von *älteren Zielpersonen ab 70 Jahren* erweisen sich von den Sample-Point bezogenen Faktoren jetzt auch *große Wohnblocks* im Vergleich zu 1-2 Familienhäusern etwas schlechter ausgeschöpft. Zumindest in den neuen Bundesländern würde dies, zusammen mit dem hier ebenfalls feststellbaren größeren Nonresponse in *Kerngebieten*, auf mögliche Viktimisierungängste, die zu einer systematisch geringeren Teilnahmebereitschaft führen, hindeuten. Da dieser Effekt jedoch erst im dritten Modell signifikant wird, ist dies vorrangig auf Wechselwirkungen mit interviewerbezogenen Faktoren zurückzuführen. In der eigentlichen Nonresponse-Analyse (Modell 2) war dieses Merkmal für sich genommen nicht signifikant. Es schlägt demnach bei der Beurteilung der Güte der Stichprobe für sich genommen nicht durch.

Bei den interviewerbezogenen Faktoren kommt vor allen Dingen die *Art des Kontaktierens* zum Tragen. Wer eher zurückhaltender ist und nicht unmittelbar und ohne Vorankündigung oder Terminabsprachen versucht, das Interview zu realisieren, ist etwas weniger erfolgreich. Der im Vergleich größere Erfolg von Interviewern, die auch *weiter von ihrem Wohnort* entfernt Befragungen durchführen, hat mit der Feldorganisation zu tun. Im Rahmen der Nachbearbeitung wurden Wechsel mit erfolgreicheren Interviewern vorgenommen, auch wenn diese ggf. dafür weitere Reisen in Kauf nehmen mussten. In den neuen Bundesländern fällt als Merkmal auch die Präferenz von *Studien mit vorgegebenen Adressen* auf. Darüber hinaus waren hier *jüngere Interviewer* unter 40 etwas weniger erfolgreich. Bezüglich des Geschlechts lässt sich für die alten und neuen Bundesländer feststellen, das *männliche Interviewer*, die weibliche Zielpersonen befragen, ebenfalls etwas schlechter abschneiden.

Alle sonstigen Merkmale, wie zum Beispiel Bildungsniveau oder auch Institutserfahrung, sind hingegen nicht signifikant. Insgesamt erhöhen die interviewerbezogenen Faktoren die Schätzstärke des Modells insbesondere in den neuen Bundesländern in einem gewissen Umfang. Dies deutet auf den vergleichsweise stärkeren Einfluss von interviewerbezogenen Faktoren als Ursache für Nonresponse hin. Im Einzelnen scheint es vor allem der Habitus, sprich die konkrete Art und insbesondere die Offensivität beim Kontakten zu

Rahmen der Nachbearbeitung, wo ggf. besonders erfolglose Interviewer durch andere ersetzt wurden. Im Falle eines Interviewerwechsels wurden von uns entweder die Merkmale des zuletzt erfolgreichen Interviewers oder aber, im Falle eines Nonresponses, die Merkmale des Interviewers, der im Rahmen der Hauptbearbeitung zuerst kontaktiert hat, zugespielt.

sein, die den Unterschied ausmacht. Die Erfahrung spielt beim ALLBUS 2000 hingegen eine eher geringere Rolle.²⁸ Alles in allem klärt jedoch auch das dritte Modell insbesondere in den alten Ländern nur relativ wenig Varianz auf.²⁹

Ausschließlich zur Veranschaulichung der Ergebnisse des Logit-Modells möchten wir eine einzelne bivariate Verteilung ausweisen: die Struktur der gültigen Brutto- und die der tatsächlich realisierten Nettostichprobe nach den Sinus-Milieus. Wie aus Übersicht 6 hervorgeht, ergeben sich zwischen Bruttovorgabe und realisierter Nettostruktur auch quantitativ keine wirklich relevanten Unterschiede. Die Abweichungen sind zwar bivariat signifikant, jedoch quantitativ wenig bedeutsam.

Übersicht 6: Verteilung MOSAIC Sinus-Milieus im ALLBUS 2000

Angaben in Prozent	Brutto ¹⁾	Netto
MOSAIC Sinus-Milieus ²⁾		
Etabliertes Milieu	8	8
Intellektuelles Milieu	9	8
Postmodernes Milieu	5	4
Modernes bürgerliches Milieu	7	7
Adaptives Milieu	6	6
Statusorientiertes Milieu	20	23
Traditionelles bürgerliches Milieu	14	13
Traditionelles Arbeitermilieu	5	4
Hedonistisches Milieu	7	8
Konsum-materialistisches Milieu	8	8
Keine Angabe	11	11

1) „Gültiges“ Stichprobenbrutto ohne qualitätsneutrale Ausfälle

2) Angaben sind hier nur für die alten Bundesländer verfügbar

Infratest Sozialforschung, eigene Berechnungen

Definitionsgemäß lassen zufallsbasierte Stichproben mit einer mathematisch berechenbaren Wahrscheinlichkeit Rückschlüsse auf die Grundgesamtheit zu. Diese Aussage darf

28 Schnell weist darauf hin, dass neben der Kontaktstrategie auch der Kontaktzeit eine größere Bedeutung zukommt. Für den anglo-amerikanischen Raum wird die Zeit zwischen 17 und 22 Uhr als beste Zeit für einen ersten Kontakt angegeben. Entsprechende Daten zum Erstkontakt beim ALLBUS 2000 liegen uns allerdings nicht vor, so dass wir dieses Merkmal nicht mit einbeziehen können (Schnell 1997: 228).

29 Dieser Befund deckt sich mit der Interpretation von Schnell, nach dessen Ansicht es bisher nicht gelungen sei, fixe Interviewermerkmale erfolgreich zur Prognose von Ausschöpfungsquoten zu verwenden (Schnell 1997: 195).

jedoch nicht so interpretiert werden, dass man zum Beispiel dann, wenn man eine ganze Reihe von Zufallsstichproben parallel ziehen würde, *ceteris paribus* keine signifikanten Einzelabweichungen bekäme. Insbesondere bei komplexen Stichprobendesigns, wie sie heute in den Sozialwissenschaften üblich sind, ist vielmehr das Gegenteil hochwahrscheinlich, ohne dass damit hinreichend sicher geklärt wäre, dass die inkriminierte Stichprobe in ihren Ergebnissen auch insgesamt „schlechter“ sei. Entscheidend ist nicht, ob einzelne Abweichungen von einer definierten Referenz zum Beispiel nach dem χ^2 -Test, signifikant sind. Es kommt vielmehr darauf an, zu prüfen, ob die Stichprobe ein systematisches Verzerrungsmuster aufweist und – immer vorausgesetzt, dass mögliche einzelne Abweichungen nicht aus zum Beispiel formalen Gründen als nicht mehr hinnehmbar betrachtet werden – wie bedeutsam die Verzerrungen für die zu untersuchenden Fragestellungen sind. Genau aus diesem Grund haben wir eine multivariate Nonresponse-Analyse anhand von Logit-Modellschätzungen durchgeführt, da sich nur auf diese Weise signifikante und bedeutsame Muster eines möglichen „nonresponse bias“ im Zusammenhang überprüfen lassen.³⁰

Auf Basis der durchgeführten Nonresponse-Analyse beim ALLBUS 2000 ergeben sich für die alten Länder – trotz der „geringen“ Ausschöpfung von nur 46,9 Prozent – so gut wie keine und für die neuen Länder – bei einer etwas besseren Ausschöpfung von 53,7 Prozent – nur schwache Anhaltspunkte dafür, dass die eingangs dargestellten Effekte von Nonresponse in irgendeiner relevanten Weise zu systematischen Verzerrungen geführt hätten. Die Gruppen, die üblicherweise einen höheren Nonresponse aufweisen, sind im ALLBUS 2000 hinreichend repräsentiert.

5. Wirkungen der Nachbearbeitung

Eines der Spezifika des ALLBUS-Verfahrensmodells der Jahre 1994, 1996 und 2000 besteht darin, dass im Anschluss an die eigentliche Hauptbearbeitungswelle, die Adressen, die nicht erfolgreich ausgeschöpft werden konnten, in mehreren Wellen nachbearbeitet wurden. Die Nachbearbeitung stützt sich demnach im Grundsatz auf die gleichen Adressen wie die Hauptbearbeitung.³¹

30 Vergleiche hierzu insbesondere Rendtel (1995: 198ff.), der diesen Zugang exemplarisch am Beispiel des Sozio-ökonomischen Panels begründet und angewendet hat.

31 Eine Ausnahme stellt der im Rahmen der Nachbearbeitung vorgenommene Austausch von solchen Befragungsklumpen dar, bei denen in der Hauptbearbeitung aus den 10 vorgegebenen Adressen gar kein oder nur ein Interview realisiert werden konnte. In diesem Fall wurde ein Reserveklumpen eingesetzt. In etwa 15 Prozent der Adressen der Nachbearbeitung wurden neu eingesetzt. 85% waren hingegen bereits im Rahmen der Hauptbearbeitung kontaktiert worden.

Es ist vor allem die Intensität der Nachbearbeitung, die zu dem von Koch (2002) dargestellten Anstieg der Fallpreise von 170 DM beim ALLBUS 1994 auf 250 DM beim ALLBUS 2000 geführt hat.³² Um die Frage der Effektivität der Nachbearbeitung unter qualitativen Gesichtspunkten beurteilen zu können, möchten wir unser Nonresponse-Analysemodell auf den Vergleich der Verteilung der Hauptbearbeitung mit den Effekten der Nachbearbeitung ausdehnen. Testhypothese ist hierbei, dass sich die Verteilung der Hauptbearbeitungswelle signifikant und entsprechend stark anhand der für Nonresponse charakteristischen Merkmale von der im Rahmen der Nachbearbeitung realisierten Verteilung unterscheidet.

5.1 Operationalisierung

Die Operationalisierung der Merkmale erfolgt in diesem Fall anhand von ausgewählten Variablen, die in den ALLBUS-Untersuchungen 1994, 1996 und 2000 gemeinsam verfügbar sind (vergleiche dazu Übersicht 7).

Berechnet wurde für jede der drei genannten ALLBUS-Querschnittserhebungen anhand der jeweils gleichen Variablen ein eigenständiges Logit-Modell. Die Ergebnisdarstellung erfolgt wiederum genauso, wie vorher bei der eigentlichen Nonresponse-Analyse. Ein negatives Vorzeichen (-) bei den Regressionskoeffizienten deutet auf eine stärkere Realisierung im Rahmen der Nachbearbeitung hin. Maßgeblich sind natürlich auch hier nur signifikante Zusammenhänge. Bei der Interpretation der Ergebnisse ist neben der Beurteilung der Stärke des Schätzmodells in diesem Fall auch von Bedeutung, inwieweit sich die möglichen Effekte konstant bei allen 3 ALLBUS-Untersuchungen vergleichbar ausprägen. Die Anlage der Nachbearbeitung als Neukontaktierung bereits vom Interviewer angegangener Adressen setzt ja geradezu voraus, dass sich die Wellen systematisch unterscheiden. Die eigentlich interessante Frage ist, inwieweit sich anhand der Nachbearbeitung bei allen drei ALLBUS-Untersuchungen mit Registerstichprobe ein einheitliches Muster bezüglich solcher Gruppen aufzeigen lässt, die bei einer „normalen“ Bearbeitung schlechter befragbar und damit anfälliger für systematischen Nonresponse sein müssten.

32 Eine weitere Ursache für den deutlich höheren Fallpreis des Jahres 2000 ergibt sich auch dadurch, dass in diesem Fall aufgrund der Umstellung von PAPI (paper pencil assisted personal interviewing) auf CAPI (computer assisted personal interviewing) eine zusätzliche Methodenstudie mit parallelisierten CAPI und PAPI-Interviews pro Point durchgeführt wurde.

Übersicht 7: Wirkungen und Effekte der Nachbearbeitung der ALLBUS-Untersuchungen – Operationalisierung anhand von geeigneten ALLBUS-Variablen (ALLBUS 94, 96, 2000)

Testhypothese:

Die Struktur der Ergebnisse der Hauptbearbeitung unterscheidet sich in einer oder mehrerer der folgenden Dimensionen signifikant und hinreichend stark von den Ergebnissen der Nachbearbeitung

1. <i>Erreichbarkeit</i>	
Haushaltsgröße	Single-Haushalt versus größere Haushalte (2 bzw. 3 Pers. u.m.)
Familienstand	Alleinstehend versus Verheiratet
(Erwerbs-)Status	Vollzeit-Erwerbstätig/Studierende versus TZ-/Nicht Erwerbstätig
BIK-Typ	Kerngebiet versus Randlage / Peripherie / ländlicher Raum
2. <i>Gebrechlichkeit / Krankheit</i>	
Alter	Befragte ab 70 Jahre versus sonstige Altersgruppen
3. <i>Misstrauen aufgrund von Viktimisierungssängsten</i>	
BIK-Typ	Kerngebiet versus Randlage/Peripherie/ländlicher Raum
Wohnstatus	Soz. Wohnungsbau versus Mieter/Eigentümer
4. <i>„Belastung“ bzw. kritische Distanz gegenüber Umfrageforschung (Misstrauen/Schutz der Privatsphäre)</i>	
Links/Rechts	Stark Links / Rechts versus Mitte/eher Links/eher Rechts
Inglehart-Index	Postmaterialisten versus Materialisten/Mischformen
Parteienpräferenz	Verweigerung / Sonstige versus „etablierte“ Parteien
5. <i>Fehlendes Interesse und / oder generelle soziale Distanz</i>	
Soziale Schicht	Unterschicht versus Mittel-/Oberschicht
Pers. wirtschaftliche Lage	Schlecht versus Durchschnitt/Gut
Höchster Schulabschluss	Hauptschule versus Abitur/Mittlere Reife
Interesse an Politik	Gering versus Mittel/Hoch

Infratest Sozialforschung, eigene Darstellung

5.2. Ergebnisse

Die „Extensität“ der Nachbearbeitung hat bei den ALLBUS-Erhebungen von Erhebungsjahr zu Erhebungsjahr zugenommen. Auch hier sticht wieder der ALLBUS 2000 mit einer nach Abschluss der Hauptbearbeitungsphase realisierten Ausschöpfung von nur 30,4 Prozent in den alten und 36,9 Prozent in den neuen Bundesländern, hervor.

Umso erstaunlicher ist es, dass sich in der Modellanalyse insbesondere für die alten Bundesländer wiederum nur sehr schwache Zusammenhänge zwischen den erwartbaren Effekten von systematischem Nonresponse und Realisierung der Interviews im Rahmen der Haupt- oder der Nachbearbeitung zeigen (vergleiche dazu Übersicht 8). Unabhängig von der Höhe der Ausschöpfung erweisen sich die Logit-Modelle für die Jahre 1994, 1996 und 2000 mit einem Pseudo-R² jeweils um die 0,03 als äußerst schwach und wenig trennscharf.

In den neuen Bundesländern ist die Güte der Modelle interessanterweise etwas größer. Bemerkenswerterweise wird in diesem Fall für das Erhebungsjahr 1994, also dem Jahr, in dem nicht mehr als 7 Prozent der Interviews durch die Nachbearbeitung dazu kamen und mit 52,2 Prozent bereits nach Abschluss der Hauptbearbeitung eine Ausschöpfung im Rahmen des damals Erwartbaren realisiert wurde, mit einem Pseudo- R^2 von 0,12 eine erstmals akzeptable Größe erreicht, die auf einen hinreichend starken systematischen Unterschied zwischen Haupt- und Nachbearbeitung hindeutet. Auch in den neuen Bundesländern hat die Bedeutung der Nachbearbeitung für die Qualität der Stichprobe seitdem kontinuierlich und umgekehrt proportional zum gestiegenen Aufwand abgenommen. Beim ALLBUS 2000 findet sich auch für die neuen Länder ein Pseudo- R^2 von nur noch 0,07.

Betrachtet man die Effekte im Einzelnen, so zeigt sich, dass durch die Nachbearbeitung über die Jahre hinweg vergleichbare Verbesserungen ausschließlich im Bereich der BIK-Regionen, sprich bezüglich der Untererfassung von Zielpersonen aus Kerngebieten, erreichbar waren. Dieser Effekt ist allerdings zum Teil auch als Artefakt der Anlage der Nachbearbeitung geschuldet, da es sich bei den Klumpen, die aufgrund mangelhafter Bearbeitung im Rahmen der Nachbearbeitung ausgetauscht und neu in die Stichprobe aufgenommen wurden, vorrangig um solche aus Kerngebieten gehandelt hat. Alle anderen signifikanten Effekte sind hingegen eher uneinheitlich und auf einzelne Erhebungsjahre begrenzt.

Besonders interessant ist, dass die bei der Nonresponse-Analyse festgestellten systematischen Abweichungen beim Alter – sprich die Untererfassung von älteren Zielpersonen ab 70 Jahren – durch die Nachbearbeitung nicht abgemildert wurden. Diesbezüglich hat sich am Ergebnis so gut wie nichts geändert. Ebenfalls nicht signifikant sind die Indikatoren Haushaltsgröße / Single-Haushalte bzw. Erwerbsstatus / Vollzeit-Erwerbstätig. Gemeinhin schlechter erreichbare Gruppen sind durch die Nachbearbeitung in der Stichprobe nicht besser repräsentiert worden. Die Variable Wohntyp / Sozialer Wohnungsbau als möglicher Indikator für Misstrauen und Viktimierungsängste ist über die Jahre hinweg ebenfalls nicht signifikant. Beim ALLBUS 1996 wurden in der Hauptidehebung sogar geringfügig mehr Zielpersonen aus dem sozialen Wohnungsbau erreicht als in normalen Mietwohnungen. Auch die Variablen, die einen möglichen Nonresponse bei postmaterialistisch orientierten oder aber aus sonstigen Gründen gegenüber Befragungen misstrauischen Personen abbilden können, sind uneinheitlich. Der „Inglehart-Index“ weist beim ALLBUS 2000 im Rahmen der Nachbearbeitung auf eine geringfügig höhere Befragung von Materialisten im Vergleich zu Postmaterialisten hin. Stabil ist dieses Muster über die Jahre hinweg jedoch nicht gewesen. Die z. T. auch signifikanten Ausprägungen bei der Links-Rechts-Positionierung und bei den Ergebnissen der Sonntagsfrage deuten ebenfalls nicht auf eine Unterrepräsentierung von gegenüber dem Mainstream eher „extremere“ Positionen hin.

Übersicht 8: Unterschiede zwischen Haupt- und Nachbearbeitungsphase bei den ALLBUS-Untersuchungen 1994, 1996 und 2000

Logistische Regressionsanalyse: Nach- versus Hauptbearbeitung

	ALLBUS 2000		ALLBUS 1996		ALLBUS 1994	
	West	Ost	West	Ost	West	Ost
Fallzahl ¹⁾ :	2.036	1.102	2.402	1.116	2.342	1.108
Ausschöpfung insgesamt	46,9	53,7	54,2	54,2	53,2	55,2
Ausschöpfung Hauptbearbeitung	30,4	36,9	41,7	46,1	44,6	52,2
Fallzahlen (in Prozent)						
Hauptbearbeitung	63	66	72	78	79	93
Nachbearbeitung	37	34	28	22	21	7
Prädiktoren						
Haushaltsgröße						
Ref.-Kat: Single						
2 Personen	-,069	,205	-,071	,345	,062	,625
3 u.m. Personen	,088	,469	,225	,609	,395*	,406
Familienstand						
Alleinstehend versus						
Verheiratet	-,102	,144	-,221	,078	,172	-,183
Erwerbsstatus						
Ref.-Kat.: Vollzeit/Student						
Teilzeit	-,182	-,094	-,114	,426	,272	,536
Rentner/Pensionär	-,010	,259	-,218	,640	-,053	,047
Arbeitslos	,105	-,037	-,010	,313	,118	,406
Sonstiges	-,222	,177	,142	,075	-,102	-,178
BIK-Typ						
Ref.-Kat.: Kerngebiet						
Rand von Stadtregionen	,426**	1,056**	,123	-,637*	-,029	1,896
Peripherie	,447**	,804**	,422**	,047	,215	-,494
Ländlicher Raum	,693**	,134	,335**	,466*	,165	,966*
Alter						
Ref.-Kat.: 70 und älter						
18-29	,102	-,112	-,291	,273	,283	,151
30-39	-,153	-,080	-,491	,032	,069	,112
40-49	-,252	-,426	-,279	,309	,181	,135
50-59	,076	-,150	-,409	,048	,180	,158
60-69	-,149	,546*	-,053	,090	-,116	,403
Wohnstatus						
Ref.-Kat.: Sozialer Wohnungsbau						
Sonstige Mietwohnung	-,152	,138	-,386*	-,593**	-,004	,062
Eigentum	-,123	,377	-,166	-,036	-,087	,642
Schicht						
Ref.-Kat.: Untere Schichten						
Mittelschicht	,065	,233	,121	,298	,047	,191
Obere Schichten	,252	1,145*	,330	-,401	-,046	1,246

Weiß nicht	-,276	-,075	-,056	135	-,560*	-1,099*
Fortsetzung der Regressionsanalyse:						
	ALLBUS 2000		ALLBUS 1996		ALLBUS 1994	
	West	Ost	West	Ost	West	Ost
Persönliche wirtschaftliche Lage						
Ref.-Kat.: (Eher) Schlecht						
Teil/Teils	-,206	-,278	-,021	,640*	-,163	,484
(Eher) gut	-,313	-,268	-,236	,756**	-,249	,404
Höchster Schulabschluss						
Ref.-Kat.: Volks-/Hauptschule						
Mittlere Reife	-,060	-,333	-,223	,082	-,078	-,035
FH-Reife	-,126	-,233	-,075	,086	-,201	1,241
Abitur	-,110	-,041	-,219	,483	-,170	-,304
Sonstiges	,093	-,200	,238	,731	,548	-,893
Interesse an Politik						
Ref.-Kat.: Wenig						
Stark	,188	-,275	,264	,221	-,009	,499
Mittel	,039	-,267	,151	,374*	-,156	-,118
Links-Rechts Positionierung						
Ref.-Kat.: Extrem Links/Rechts						
Eher Links	,217	-,060	,066	-,551	-,484	,266
Mitte	,153	-,373	-,066	-,427	-,675**	,493
Eher Rechts	,246	-,522	,111	-,283	-,540*	,752
Keine Angabe	,315	-,200	-,897	-1,902*	-,432	,169
Inglehart-Index						
Ref.-Kat.: Postmaterialisten						
PM-Mischtyp	-,031	-,311	-,068	,659*	,196	,504
M-Mischtyp	-,043	-,237	-,036	,564*	-,016	,794
Materialisten	-,328*	-,572*	,169	,481	,278	,774
Sonntagsfrage						
Ref.-Kat.: Angabe verw./Sonst.						
CDU/CSU	,219	-,572*	-,206	-,669*	,063	,465
SPD	,281	-,433*	,014	-,178	,010	-,254
FDP	,311	-,299	-,441	-,662	,069	,650
Grüne	,464	-,177	-,177	-,851	-,120	,346
Würde nicht wählen	,608*	-,163	,108	-,452	,247	,489
Weiß nicht	,493*	,017	,063	-,396	-,113	,170
Modellkonstante	,183	,920	1,197*	-,028	1,646**	,255
Pseudo R² (Mc-Fadden)	0,028	0,068	0,032	0,091	0,029	0,124
1) Fallzahlen der ALLBUS-Erhebungen. Bedingt durch fehlende Angaben bei den einzelnen Merkmalen wurden die Modelle mit einer geringfügig kleineren Fallzahl berechnet						
Binär-logistische Regression (SPSS 11.0 - NomReg) : Nachbearbeitung (0) versus Hauptbearbeitung (1)						
B-Koeffizienten mit * = Sig.-Niveau < 0,05 bzw. mit ** = Sig.-Niveau < 0,01						

Zur Veranschaulichung möchten wir auch in diesem Fall wieder einzelne bivariate Verteilungen darstellen. Ausgewählt haben wir die Ergebnisse der Rechts-Links-Positionierung sowie den Inglehart-Index im Vergleich Hauptbearbeitung und Endergebnis.

Übersicht 9: Verteilung der Rechts-Links Einstufung und des Inglehart-Index nach der Hauptbearbeitungswelle und im Endergebnis beim ALLBUS 2000

Angaben in Prozent	West		Ost	
	Haupt-bearbeitung	End-ergebnis	Haupt-bearbeitung	End-ergebnis
Rechts-Links Einstufung				
Links (1-2)	4	4	10	9
Eher links (3-4)	24	25	32	30
Mitte (5-6)	40	40	42	44
Eher rechts (7-8)	18	17	7	8
Rechts (9-10)	2	2	1	1
Keine Angabe	12	12	8	8
Inglehart-Index				
Postmaterialisten (P)	28	28	17	16
PM-Mischtyp	28	27	33	34
M-Mischtyp	31	30	33	32
Materialisten (M)	12	13	16	17
Keine Angabe	1	2	1	1
1) Zur Indexbildung vergleiche Codebuch ALLBUS 2000 (ZA-Nr. 3451), Note 2: 494				

Infratest Sozialforschung, eigene Berechnungen

Das Ergebnis nach der Hauptbearbeitung unterscheidet sich fast nicht vom Endergebnis.³³ In der Interpretation ergäbe sich kaum ein relevanter Unterschied, wenn man stattdessen nur auf die im Rahmen der Hauptbearbeitung realisierten Fälle zurückgreifen würde.³⁴ Die durchgeführte Analyse zeigt, dass die Abweichungen bei den drei ALLBUS-Untersuchungen von 1994, 1996 und 2000 nicht systematisch verteilt sind. Ein einheitli-

33 Zur relativ geringen Bedeutung von möglichen bivariaten Abweichungen vergleiche die hierzu bereits im letzten Abschnitt gemachten grundsätzlichen Anmerkungen.

34 Es lässt sich übrigens schwer abschätzen, welche Verteilung „wahrer“ ist. Interessanterweise ist die Abweichung im Osten etwas stärker ausgeprägt. Hier hatten wir zwar eine etwas höhere Ausschöpfung, aber trotzdem auch eine leicht stärkere Abweichungssystematik in der Verteilung des Nonresponses. Auch der Vergleich der Haupt- mit der Nachbearbeitung ergab eine leicht stärkere Abweichungssystematik als in den alten Ländern.

ches Muster ist – mit Ausnahme der Verbesserung der Ausschöpfung bei den BIK-Kerngebieten – nicht zu erkennen. Insgesamt gibt es bedingt durch die Nachbearbeitung Effekte. Art und Richtung der Effekte lassen sich jedoch weder hinreichend gut vorhersagen, noch in einen klaren Zusammenhang zu einer besseren Stichprobenqualität bringen. Durch die intensive Nachbearbeitung hat sich demnach vorrangig die Fallzahl, also die Quantität, und weniger die Qualität erhöht. Natürlich ergeben sich in den Strukturen der Haupt- und der Nachbearbeitung gewisse Abweichungen. Diese Unterschiede sind im Vergleich über die Erhebungsjahre jedoch kaum systematisch und für sich genommen relativ gering ausgeprägt. Bemerkenswerterweise sind die Abweichungen auch in diesem Fall in den neuen Ländern, trotz höherer Stichprobenausschöpfung, größer als in den alten Ländern.

6. Fazit und Diskussion

Die 1994, 1996 und 2000 durchgeführten, hart bearbeiteten und dokumentierten registerbasierten Verfahrensmodelle bei den ALLBUS-Erhebungen (best practice) gingen mit einem weiteren Rückgang in der Ausschöpfung – zuletzt beim ALLBUS 2000 auf 46,9 Prozent für die alten und 53,7 Prozent für die neuen Bundesländer – einher. In der vorliegenden Arbeit haben wir die möglichen Qualitätseffekte anhand von drei verschiedenen Zugängen untersucht.

Die im ersten Schritt vorgenommene **Nettovalidierung** der Stichproben am Mikrozensus als Referenzstatistik erbrachte keinerlei neue Hinweise auf mögliche gravierende Verzerrungen. Vielmehr ergab sich eine akzeptable Übereinstimmung der soziodemografischen Randverteilungen mit der amtlichen Statistik. Im Einzelnen erwiesen sich die relativ schlechter ausgeschöpften ALLBUS-Erhebungen in den alten Bundesländern – hierbei insbesondere der ALLBUS 2000 – im Vergleich sogar tendenziell als noch besser angepasst.

Auch die im zweiten Schritt durchgeführte **multivariate Nonresponse-Analyse** von verfügbaren Prädiktoren, die zu einem auch theoretisch begründeten „nonresponse bias“ führen könnten, hat die Hypothese nicht bestätigt, dass die sinkenden Ausschöpfungsraten beim ALLBUS mit einer zunehmenden systematischen Verzerrung einhergehen. Hierbei wurden zusätzlich, neben den von den Einwohnermeldeämtern im Rahmen der Stichprobenziehung mitgelieferten demografischen Informationen, externe und kommerziell vertriebene mikrogeografische Wohnumfeldindikatoren (MOSAIC-Variablen und -Milieudaten) herangezogen. Im Einzelnen ließ sich die für allgemeine Bevölkerungsumfragen mehr oder weniger übliche Unterrepräsentierung von Seniorinnen und Senioren (70 Jahre und älter) feststellen, die vor allem aufgrund der im höheren Lebensalter verstärkt auftretenden Probleme einer Hilfe- oder Pflegebedürftigkeit schwerer zu befragen

sind. Eine etwas stärkere Systematik in der Verteilung des Nonresponses ergab sich – interessanterweise vorrangig bei der relativ besser ausgeschöpften Stichprobe für die neuen Bundesländer – erst bei Einbeziehung weiterer interviewerbezogener Merkmale. Als signifikant erwies sich für die neuen Länder nun auch zusätzlich eine tendenzielle Untererfassung von Zielpersonen aus Kerngebieten von Stadtregionen bzw. aus größeren Wohnblocks oder Hochhäusern. Dieser Effekt deutet für die neuen Länder auf offenbar etwas stärker wirksame mögliche Viktimisierungängste hin, die die Kooperationsbereitschaft von Zielpersonen bei face-to-face Befragungen im Haushalt einschränken können. Bei den interviewerbezogenen Faktoren erwiesen sich die Merkmale als signifikant, die mit dem „Habitus“ bzw. der Art des Kontaktierens zusammenhängen. Tendenziell erfolgreicher wurden die Points bearbeitet, für die Interviewer eingesetzt werden konnten, die sehr offensiv (zum Beispiel ohne persönliche Vorankündigung oder Terminabsprachen) agieren. Alles in allem ist der Einfluss auch dieser Faktoren auf die Gesamtverteilung des Nonresponses jedoch ebenfalls gering.

Auch die im dritten Schritt vorgenommene **Analyse der Effektivität** der für harte Verfahrensmodelle typischen **intensiven Nachbearbeitung** bestätigte den Gesamtbefund. Es ergaben sich im Vergleich der drei ALLBUS-Untersuchungen keinerlei wirklich aussagekräftige Hinweise darauf, dass die realisierten Nettostichproben trotz der hohen Nonresponse-Raten mit einer systematischen Verzerrung verbunden waren, die erst durch die Nachbearbeitung tendenziell ausgeglichen werden konnten. Im Gegenteil ergab die hierbei analog angewandte multivariate Analyse für die alten Länder so gut wie gar keine, und für die neuen Länder in der Tendenz nur schwache Hinweise auf eine systematisch und nicht zufällig andere Verteilung innerhalb der Haupt- und der Nachbearbeitung. Da die Nachbearbeitung weitestgehend auf den gleichen Fällen basierte, die bereits im Rahmen der Hauptbearbeitung kontaktiert wurden, ohne dabei erfolgreich befragt worden zu sein, ist dies genau das Gegenteil vom dem, was eigentlich intendiert war. Tatsächlich hat die zeit- und auch außerordentlich kostenintensive Nachbearbeitung vorrangig Quantität und weniger Qualität (im Sinne einer Reduzierung möglicher Verzerrungen) gebracht. Das Ergebnis ist eine etwas höhere Ausschöpfungsrate, die letztendlich nur zu einer etwas höheren Fallzahl führt. Letzteres ließe sich, bei deutlich geringerem Aufwand, auch dadurch erreichen, dass die Bruttostichprobe der Hauptbearbeitungswelle von vornherein etwas größer angelegt wird.

Zusammengefasst bestätigen die Analysen, dass die Ausschöpfung für sich genommen kein hinreichend geeigneter Indikator zur Beurteilung der Qualität einer Bevölkerungsumfrage ist. Die Ergebnisse zeigen vielmehr, dass auch „schlecht“ ausgeschöpfte Stichproben offenbar gute Qualität aufweisen können. Wir interpretieren dies als überzeugenden Beleg für die prinzipielle Leistungsfähigkeit von zufallsbasierten Stichprobenverfah-

ren. Letztendlich ist es diesem fundamentalen Prinzip geschuldet, die Auswahl der Zielpersonen streng zufällig und losgelöst von möglichen gezielten Suchstrategien der Interviewerinnen und Interviewer vorzunehmen, dass sich auch der Stichproben-Nonresponse mehr oder weniger zufällig verteilt und nicht zu relevanten systematischen Verzerrungen führt. Voraussetzung hierfür ist, dass die Stichprobe „kunstgerecht“ und professionell gezogen wurde und dass auch die Bearbeitung im Feld sorgfältig genug und mit der hierfür notwendigen Erfahrung vorgenommen wird. Die Ergebnisse unserer Analyse zu den Effekten der Nachbearbeitung zeigen allerdings, dass es hierbei nicht unbedingt darauf ankommt, den Aufwand in der Feldarbeit immer weiter zu erhöhen. Das „Geheimnis“ des Erfolgs liegt offenbar in der soliden Bearbeitung in der Breite, und weniger darin, in diversen Nachbearbeitungswellen immer wieder aufs Neue bestimmte schlecht ausgeschöpfte Sample-Points anzugehen.

In der Konsequenz bedeutet dies, dass im Rahmen der notwendigen Qualitätsdebatte die Bedeutung des Merkmals Stichprobenausschöpfung deutlich relativiert werden sollte. Es ist nicht nur so, dass es keinen Sinn macht, Ausschöpfungen zu vergleichen, ohne die dahinter stehende Art der Definition eines Stichprobenbruttos zu überprüfen. Auch grundsätzlich ist dieses Merkmal für sich genommen als Prädiktor zur Beurteilung der Ergebnisqualität von vergleichbar angelegten Zufallsstichproben wenig aussagekräftig. Aus unserer Sicht wäre es sinnvoller, stattdessen bereits bei der Anlage der Untersuchung mit zu klären, anhand welcher externer Kriterien die Stichprobe validiert werden kann.³⁵ Dies können zum Beispiel die soziodemografischen Verteilungen der amtlichen Statistik sein. Dies kann aber auch anhand von Referenzinstrumenten geschehen, die in entsprechender Form auch in anderen Erhebungen eingesetzt werden. Hilfreich wäre, wenn zumindest größere allgemeine Bevölkerungsumfragen einen bestimmten Apparat von vergleichbaren Indikatoren einsetzen würden, die einerseits in einem darstellbaren Zusammenhang zum Untersuchungsziel (sprich der „abhängigen Variable“) stehen und die andererseits als allgemeine Referenz für andere Erhebungen nutzbar wären. Damit wäre die Möglichkeit eines gezielten Monitorings von repräsentativ angelegten Bevölkerungsstichproben eröffnet. Nonresponse-Studien werden hingegen, aufgrund der damit verbundenen umfangreichen methodischen Voraussetzungen, auch in Zukunft wohl eher die Ausnahme bleiben. Das von uns eingeführte Modell einer (anonymisierten) Verknüpfung von Stichprobendaten und externen Informationen erscheint uns hierfür jedoch hilfreich und weiterführend zu sein.

35 Vergleiche dazu auch die Überlegungen von Christian von der Heyde (von der Heyde 1998).

Abschließend möchten wir noch einmal darauf verweisen, dass die Ergebnisse unserer Untersuchung auf der Re-Analyse der einwohnermeldeamtsbasierten ALLBUS-Erhebungen beruhen. Tatsächlich ist dieses best-practice Verfahrensmodell für die Breite der empirischen Sozialforschung nicht typisch und aufgrund des damit verbundenen Zeit- und Kostenaufwandes auch nicht als grundsätzlicher Standard vorzugeben. Im Bereich der persönlich-mündlichen Erhebungen sind als alternative Modelle verschiedene Formen von Random-Route Verfahren (ADM-Design) eingeführt, die ebenfalls auf einer Zufallsauswahl beruhen, jedoch aufgrund der Integration von Stichprobenbildung und Befragung deutlich forschungsökonomischer realisiert werden können.³⁶ Im Vergleich zum best-practice Modell sind Random-Route Verfahren pragmatischer, insbesondere was die Kontrollierbarkeit der Vorgaben sowie die Intensität, mit der ein Sample-Point ausgeschöpft wird, angeht. Andererseits unterscheiden sie sich jedoch fundamental durch ihren grundsätzlich zufallsbasierten Charakter von unregelmäßigen Verfahren oder auch von Quotenstichproben.³⁷ Unsere Hypothese hierzu lautet, dass auch diese Verfahren, sofern sie ein dokumentiertes Maß an Standards, zum Beispiel hinsichtlich der Zufallsverteilung der Startadressen für den Random-Walk, der Einhaltung der Begehungsvorschriften etc., aufweisen und von einem erfahrenen Interviewerstab bearbeitet werden, zu akzeptabler Qualität führen. Aus unserer Sicht macht es allerdings wenig Sinn, die „Güte“ wie bisher durch überdimensional hohe Ausschöpfungsraten darzustellen, die im Prinzip mit den Ergebnissen von härter angelegten Verfahren nicht mehr vergleichbar sind. Hier ist aus unserer Sicht mehr Mut zum Realismus gefragt.

Wir möchten uns deshalb dafür aussprechen, den faktischen „Pluralismus der Verfahrensmodelle“ auch in Zukunft beizubehalten. Dies sollte jedoch nicht als Plädoyer für einen möglichen „Methoden-laissez faire“ verstanden werden. Unsere Empfehlung geht in diesem Fall eindeutig dahin, am Zufallscharakter der Stichprobenbildung (probability sample) sowie an den hierbei erforderlichen Standards der Dokumentation festzuhalten

36 Typischerweise wird bei den heute gängigen Formen von Random-Route Erhebungen darauf verzichtet, die Adressen in einem Vorlauf erst dokumentieren zu lassen und diese danach den Interviewern extern vorzugeben (Adressenvorlauf). Die Integration von Stichprobenbildung und Befragung im Rahmen eines einheitlichen Random-Walks führt in diesem Fall zu spürbaren Kostenvorteilen, ohne dass der Charakter einer Zufallsauswahl dadurch grundsätzlich in Frage gestellt wäre.

37 Zu den Problemen von Quotenerhebungen vergleiche zum Beispiel die nach wie vor äußerst instruktive empirische Analyse von Hoag (1986): „Der Bekanntenkreis als Universum: Das Quoten-Verfahren der Shell-Studie“.

und auf Quoten-Erhebungen oder sonstige Verfahren nur in begründeten besonderen Fällen zurückzugreifen.³⁸

Korrespondenzadressen

Ulrich Schneekloth
Infratest Sozialforschung
Landsberger Straße 338
80687 München
email: Ulrich.Schneekloth@nfoeurope.com

Ingo Leven
Infratest Sozialforschung
Landsberger Straße 338
80687 München
email: Ingo.Leven@nfoeurope.com

Literatur

Andreß, H.J./Hagenaars, J.A./Kühnel, S., 1997: Analyse von Tabellen und kategorialen Daten. Berlin.

Deutsche Forschungsgemeinschaft (DFG), 1999: Qualitätskriterien der Umfrageforschung. Berlin.

Esser, H./Grohmann, H./Müller, W./Schäffer, K.-A., 1989: Mikrozensus im Wandel. Wiesbaden.

Groves, R.M., 1989: Survey Errors and Survey Costs. New York.

Hartmann, P.H., 1990: Wie repräsentativ sind Bevölkerungsumfragen. Ein Vergleich des ALLBUS und des Mikrozensus. ZUMA-Nachrichten 26: 7-30.

Hartmann, P.H./Schimpl-Neimanns, B., 1992: Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage. Kölner Zeitschrift für Soziologie und Sozialpsychologie 49: 315-340.

Heyde von der, Chr., 1998: Qualitätsmaße und Qualitätskosten. München.

38 Quoten-Erhebungen sind aus unserer Sicht bei allgemeinen Bevölkerungsumfragen grundsätzlich dann eine Alternative, wenn die Grundgesamtheit nur eine sehr kleine Teilgruppe der Bevölkerung umfasst, die bei zufallsbasierten Erhebungen nur mit sehr hohem Risiko und entsprechend hohen Kosten abgebildet werden können. Weitere Gründe stellen natürlich auch externe Rahmenbedingungen dar, zum Beispiel wenn nur sehr wenig Feldzeit für die Durchführung der Erhebung eingeräumt werden kann. Dies ändert jedoch nichts an dem grundsätzlichen Dilemma, dass bei einer quotenbasierten Erhebung auch bei optimaler Aussteuerung der vorgegebenen Zellen kein wissenschaftlich haltbarer Rückschluss zulässig ist, ob damit auch die nicht vorgegebenen Merkmalsverteilungen richtig abgebildet sind.

Hoag, W., 1986: Der Bekanntenkreis als Universum: Das Quotenverfahren der Shell-Studie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 123-132.

Koch, A./Gabler, S./Braun, M., 1994: Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1994. ZUMA-Arbeitsbericht 1994/11. Mannheim.

Koch, A., 1997: Teilnahmeverhalten beim ALLBUS 1994. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49: 98-122.

Koch, A., 1998: Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. *ZUMA-Nachrichten* 42: 66-93.

Koch, A./Wasmer, M./Harkness, J./Scholz, E., 2001: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2000. Mannheim.

Koch, A., 2002: 20 Jahre Feldarbeit im ALLBUS: ein Blick in die Blackbox; in: *ZUMA-Nachrichten* 51: 9-37.

Microm GmbH und Sinus Sociovision GmbH, 2000: MOSAIC Informationen. Stand 2000. Neuss, Frühjahr 2000.

Rendtel, U., 1995: Lebenslagen im Wandel: Panelfälle und Panelrepräsentativität. Frankfurt a. M.

Schnell, R., 1991: Wer ist das Volk? Zur faktischen Grundgesamtheit bei allgemeinen Bevölkerungsumfragen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 1: 106-137.

Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen.

Schnell, R./Hill, P. B./Esser, E., 1999: Methoden der empirischen Sozialforschung – 6. völlig überarbeitete und erweiterte Auflage. München.

Schnell, R./Kreuter, F., 2000: Das DEFECT-Projekt: Sampling Error und Nonsampling-Errors in komplexen Bevölkerungsstudien. *ZUMA-Nachrichten* 47: 89-102.

Statistisches Bundesamt, 1998: Bevölkerung und Erwerbstätigkeit. Fachserie 1 Reihe 4.1.2. Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen. Stuttgart.

Statistisches Bundesamt, 2001a: Bevölkerung und Erwerbstätigkeit. Fachserie 1 Reihe 3. Haushalte und Familien. Stuttgart.

Statistisches Bundesamt, 2001b: Bevölkerung und Erwerbstätigkeit. Fachserie 1 Reihe 4.1.2. Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen. Stuttgart.

Wasmer, M./Koch, A./Harkness, J./Gabler, S., 1997: Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1996. ZUMA-Arbeitsbericht 1996/08. Mannheim.

Westerhoven, E. van, 1978: Coverage Non-Response: Does it pay? A study of refusers and absentees. *Journal of the Market Research Society*, 20, 4: 245-247.

Anhang

**Tabelle A1: Differenz zum Mikrozensus: Merkmal Altersklassen
(Abweichung in Prozentpunkten)**

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
20 bis 29 Jahre	19,0	-0,1	17,4	0,8	14,7	-0,2
30 bis 39 Jahre	19,6	2,2	20,5	1,4	20,9	-0,8
40 bis 49 Jahre	16,3	0,6	17,2	0,6	18,1	1,7
50 bis 59 Jahre	18,5	0,0	17,6	0,4	16,2	-0,4
60 bis 69 Jahre	13,4	0,1	13,8	-0,1	15,6	0,8
≥ 70 Jahre	13,3	-2,8	13,6	-3,3	14,6	-1,2
Ost						
20 bis 29 Jahre	16,8	-1,6	15,4	-2,7	14,5	-1,4
30 bis 39 Jahre	20,7	-1,9	20,4	-0,2	19,0	1,5
40 bis 49 Jahre	16,6	1,8	17,8	1,3	19,5	0,3
50 bis 59 Jahre	19,9	0,8	18,6	2,0	16,1	-1,0
60 bis 69 Jahre	13,9	1,9	15,1	0,2	17,0	0,6
≥ 70 Jahre	12,2	-1,0	12,6	-0,4	13,9	0,1

Mikrozensus 1994, 1996 und 2000: Statistisches Bundesamt 2001a: 240

**Tabelle A2: Differenz zum Mikrozensus: Merkmal Geschlecht
(Abweichung in Prozentpunkten)**

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Männlich	47,8	3,8	47,9	2,2	48,0	0,3
Weiblich	52,2	-3,8	52,1	-2,2	52,0	-0,3
Ost						
Männlich	47,0	0,6	47,3	0,2	47,8	-0,1
Weiblich	53,0	-0,6	52,7	-0,2	52,2	0,1

Mikrozensus 1994, 1996 und 2000: Statistisches Bundesamt 2001a: 252-254

**Tabelle A3: Differenz zum Mikrozensus: Merkmal Familienstand
(Abweichung in Prozentpunkten)**

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Ledig	22,2	-0,5	22,1	-0,5	22,0	0,2
Verheiratet	62,8	1,7	62,9	2,8	62,7	0,4
Verwitwet/geschieden	15,0	-1,2	15,0	-2,3	15,3	-0,6
Ost						
Ledig	17,4	-3,1	18,9	-5,5	22,2	-0,3
Verheiratet	65,2	3,7	63,7	5,4	60,1	-0,6
Verwitwet/geschieden	17,4	-0,6	17,4	0,1	17,7	0,9

Mikrozensus 1994, 1996 und 2000: Statistisches Bundesamt 2001a: 240

**Tabelle A4: Differenz zum Mikrozensus: Merkmal Erwerbsstatus
(Abweichung in Prozentpunkten)**

	MZ 1993	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Erwerbstätige	52,2	4,4	54,1	1,7	54,9	-1,6
Erwerbslose			4,8	-1,2	4,1	-0,9
Nichterwerbspersonen	47,8	-4,4	41,1	-0,5	41,0	2,5
Ost						
Erwerbstätige	52,9	0,2	53,2	0,5	51,5	-1,0
Erwerbslose			11,5	0,2	11,6	1,0
Nichterwerbspersonen	47,1	-0,2	35,3	-0,7	36,9	0,0

Mikrozensus 2000: Statistisches Bundesamt 2001b: 160 (West) / 292 (Ost)

Mikrozensus 1996: Statistisches Bundesamt 1998: 139 (West) / 250 (Ost),

Mikrozensus 1993: Koch 1998: 76

**Tabelle A5: Differenz zum Mikrozensus: Merkmal Schulabschluss
(Abweichung in Prozentpunkten)**

	MZ 1993	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Haupt-/Volksschulabschluss	59,6	-5,7	58,6	-7,7	54,5	-4,1
Bis Realschulabschluss	23,5	4,2	21,8	3,3	23,5	0,5
Fachhoch-/Hochschulreife ¹⁾	16,9	1,5	19,6	4,4	22,0	3,6
Ost						
Bis Realschulabschluss	86,0	-4,3	84,8	-2,4	77,5	3,8
Fachhoch-/Hochschulreife ¹⁾	13,9	4,3	15,2	2,4	22,5	-3,8

1) Beim Mikrozensus 1993 ist die Fachhochschulreife in der Kategorie „Bis Realschulabschluss“ mit enthalten.

Mikrozensus 2000: Statistisches Bundesamt 2001b: 160 (West) / 292 (Ost)

Mikrozensus 1996: Statistisches Bundesamt 1998: 142 (West), 254 (Ost)

Mikrozensus 1993: Koch 1998: 76

Tabelle A6: Differenz zum Mikrozensus: Merkmal Haushaltsgröße (Abweichung in Prozentpunkten)

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Einpersonenhaushalt	20,3	-1,5	20,9	-3,9	21,2	-3,0
Mehrpersonenhaushalt	79,7	1,5	79,1	3,9	78,8	3,0
Ost						
Einpersonenhaushalt	17,0	-2,9	17,5	-3,0	19,6	0,1
Mehrpersonenhaushalt	83,0	2,9	82,5	3,0	80,4	-0,1

Mikrozensus 1994, 1996, 2000: Statistisches Bundesamt 2001a: 251

SINUS SOCIOVISION

4.1.4 Kurzcharakteristik der Sinus-Milieus 2000 (Westdeutschland)

Gesellschaftliche Leitmilieus

■ Etabliertes Milieu (vorher: Konsens-technokratisches Milieu)	10% ➤	Die erfolgsorientierte Konsum-Elite unserer Gesellschaft mit ausgeprägten Exklusivitätsansprüchen
■ Intellektuelles Milieu (vorher: liberal-intellektuelles Milieu)	10% ➤	Die aufgeklärte, postmateriell orientierte Werte-Avantgarde unserer Gesellschaft
■ Postmodernes Milieu	6% ➤	Die extrem individualistische, „multi-optionale“ Life Style-Avantgarde unserer Gesellschaft

Moderner Mainstream

■ Adaptives Milieu (vorher: Modernes Arbeitermilieu)	8% ➤	Der gut ausgebildete, mobile und pragmatische Mainstream der jungen modernen Mitte
■ Statusorientiertes Milieu (vorher: Aufstiegsorientiertes Milieu)	18% ➤	Die beruflich und sozial aufstrebende untere Mitte – die Erfolgsinsignien unserer Konsumgesellschaft im Blick
■ Modernes bürgerliches Milieu	8% ➤	Die konventionelle neue Mitte, die nach einem harmonischen, behüteten Leben in gesicherten Verhältnissen strebt

Traditioneller Mainstream

■ Traditionelles bürgerliches Milieu (vorher: Kleinbürgerliches Milieu)	14% ➤	Die Sicherheits- und Status quo-orientierte Kriegsgeneration, die an den traditionellen Werten wie Pflicht und Ordnung festhält
■ Traditionelles Arbeitermilieu	4% ➤	Die an den Notwendigkeiten des Lebens ausgerichtete traditionelle Arbeiterkultur der Eckkneipen, Kleintierzüchter und Schützenvereine

Moderne Unterschicht

■ Konsum-materialistisches Milieu (vorher: Traditionelles Arbeitermilieu)	11% ➤	Die stark hedonistisch geprägte Unterschicht, die Anschluss halten will an die Konsum-Standards der breiten Mitte
■ Hedonistisches Milieu	12% ➤	Die unangepasste junge Unterschicht, die Spass haben will und sich den Konventionen und Verhaltenserwartungen der Leistungsgesellschaft verweigert